ARMY RESEARCH LABORATORY

Proceedings of the Third Annual U.S. Army Conference on Applied Statistics, 22–24 October 1997

Barry A. Bodt
EDITOR

Hosted by:
GEORGE MASON UNIVERSITY

Cosponsored by:
U.S. ARMY RESEARCH LABORATORY
U.S. MILITARY ACADEMY
U.S. ARMY RESEARCH OFFICE
WALTER REED ARMY INSTITUTE OF RESEARCH
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
TRADOC ANALYSIS CENTER-WSMR

ARL-SR-74 July 1998

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Abstract

The third U.S. Army Conference on Applied Statistics was hosted by George Mason University (GMU) during 22–24 October 1997 at the recently opened Johnson Center on campus. The conference was cosponsored by the U.S. Army Research Laboratory (ARL), the U.S. Army Research Office (ARO), the U.S. Military Academy (USMA), the U.S. Army Training and Doctrine Command (TRADOC) Analysis Center-White Sands Missile Range, the Walter Reed Army Institute of Research (WRAIR), and the National Institute for Standards and Technology (NIST). The U.S. Army Conference on Applied Statistics is a forum for technical papers on new developments in statistical science and on the application of existing techniques to Army problems. This document is a compilation of available papers offered at the conference.

FOREWORD

The third U.S. Army Conference on Applied Statistics was hosted by George Mason University (GMU) during 22–24 October 1997 at the recently opened Johnson Center on campus. The conference was cosponsored by the U.S. Army Research Laboratory (ARL), the U.S. Army Research Office (ARO), the U.S. Military Academy (USMA), the U.S. Army Training and Doctrine Command (TRADOC) Analysis Center-White Sands Missile Range (TRAC-WSMR), the Walter Reed Army Institute of Research (WRAIR), and the National Institute for Standards and Technology (NIST). The U.S. Army Conference on Applied Statistics is a forum for technical papers on new developments in statistical science and on the application of existing techniques to Army problems. The Army faces far-ranging challenges that encompass many topics in which probability and statistics have a contribution to make. The purpose of this conference is to promote the practice of statistics in the solution of these diverse Army problems.

The third conference was preceded by a short course, "Virtual Reality and Scientific Visualization," given by Professors Edward Wegman and Dan Carr. Several opportunities were afforded conference participants to see and work with the scientific visualization tools at GMU. Distinguished speakers spoke during invited general sessions: Prof. Thomas Hettmansperger, Penn State University; Prof. Jeffrey Birch, Virginia Tech; Prof. Lyle Ungar, University of Pennsylvania; Prof. Judea Pearl, University of California, Los Angeles; Prof. Donald Berry, Duke University; Prof. Carey Priebe, Johns Hopkins University; Prof J. David Cooke, University of Western Ontario; and Dr. Eric Lagergren, NIST. Topical methodological areas included, for example, nonparametric methods, experimental design, neural networks, fuzzy control, statistical process control, genetic algorithms, robust design, density estimation, and new advances in statistical software for data visualization. Application areas included, for example, the digitized battlefield, animation of dust behavior for incorporation in simulation models, communication among robot scouts on the battlefield, chemical discrimination, parts inventory control, camouflage effectiveness, terrain modeling, and mine-field detection. A highlight of the conference was the awarding of the Army Wilks award to William Jay Conover of Texas Tech for years of substantive contribution to statistical methodology and the practice of statistics in the Army.

The Executive Board for the conference recognizes Dr. Robert Launer, ARO, Dr. Mark Vangel, NIST, and Mr. David Webb, ARL, for assisting with conference details; Dr. Jock Grynovicki for oversite of the Army Wilks Award; Dr. Barry Bodt, ARL, for general conference administration and proceedings; and Prof. Edward Wegman, GMU, for hosting the conference and handling all local arrangements. Special thanks are due Ms. Patricia Joyce, GMU, who assisted in handling many on-site details.

INTENTIONALLY LEFT BLANK.

Table of Contents*

* This table of contents contains only the papers that appear in the Proceedings.

FINAL PROGRAM
THIRD U.S. ARMY CONFERENCE ON APPLIED STATISTICS

20–24 October, 1997

Hosted by George Mason University

Cosponsored by:

U.S. Army Research Laboratory
U.S. Army Research Office
National Institute of Standards and Technology
TRADOC Analysis Center—WSMR
U.S. Military Academy
Walter Reed Army Institute of Research

Monday, 20 October 1997

0800 - 0900 REGISTRATION (George W. Johnson Center)

0900 - 1200 TUTORIAL:

VIRTUAL REALITY AND SCIENTIFIC VISUALIZATION
Edward J. Wegman and Daniel B. Carr, George Mason University

1200 - 1330 Lunch

1330 - 1600 TUTORIAL

Tuesday, 21 October 1997

0800 - 1200 TUTORIAL

1200 - 1330 Lunch

1330 - 1600 TUTORIAL

Wednesday, 22 October 1997

0800 - 0900 REGISTRATION (George W. Johnson Center)

0900 - 0930 CALL TO ORDER:

Conference Chairman: Barry Bodt, U.S. Army Research Laboratory

Conference Host: Edward Wegman, George Mason University

OPENING REMARKS:

Jagdish Chandra, Deputy Director, Information Science and Technology Directorate, U.S. Army Research Laboratory

0930 - 1200 GENERAL SESSION I

Chair: Douglas Tang, Walter Reed Army Institute of Research

0930 - 1030 KEYNOTE ADDRESS

KRUSKAL-WALLIS, MULTIPLE COMPARISONS, AND EFRON DICE
Thomas Hettmansperger, Penn State University

1030 - 1045 PRESIDENT'S WELCOME
Alan Merten, President, George Mason University

10:45-11:15 Break

1115 - 1215 MODEL ROBUST REGRESSION: MOTIVATION, RESULTS, AND APPLICATIONS
Jeffrey Birch, Virginia Tech
James E. Mays, Virginia Commonwealth University

1215 - 1330 Lunch

1330 - 1430 CONTRIBUTED SESSION I (in parallel with Contributed Sessions II and IIa)

Chair: Robert Burge, Walter Reed Army Institute of Research

SPLIT-PLOT DESIGNS FOR EXPERIMENTS THAT AUGMENT OPERATIONAL FLIGHT TESTS WITH DIGITAL SIMULATIONS
Frank B. Gray, Air Force Operational Test and Evaluation Center

DEVELOPMENT OF INFERENCE ALGORITHMS FOR WORLDWIDE APPLICATION OF STANDARD DIGITAL TERRAIN DATA TO THE NATO MOBILITY MODEL
C. Denise Bullock and Nancy A. Renfroe, Waterways Experiment Station

Wednesday, 22 October 1997 (Continued)

1330 - 1430 CONTRIBUTED SESSION II (in parallel with Contributed Sessions I and IIa)

Chair: Deloris Testerman, Yuma Proving Ground

A METHODOLOGY FOR EVALUATING PREDICTION UNCERTAINTY
Michael D. McKay, Los Alamos National Laboratory

1330 - 1430 CONTRIBUTED SESSION IIa (in parallel with Contributed Sessions I and II)

Chair: Linda Moss, U.S. Army Research Laboratory

ROBUST ADAPTIVE BEAMFORMING FOR SPATIALLY SPREAD SOURCES
Kristine L. Bell and Harry L. Van Trees, George Mason University

DOA ESTIMATION WITH HEXAGONAL ARRAYS
Zhi Tian and Harry L. Van Trees, George Mason University

1430 - 1445 Break

1445 - 1700 GENERAL SESSION II

Chair: Mark Vangel, National Institute of Standards and Technology

1445 - 1545 BUILDING RELIABLE, INTERPRETABLE MODELS USING NEURAL NETWORKS
Lyle H. Ungar, University of Pennsylvania

1545 - 1600 Break

1600 - 1700 SIMPSON'S PARADOX AND ITS PRACTICAL IMPLICATIONS
Judea Pearl, University of California, Los Angeles

1830 - WILKS AWARD BANQUET (Performing Arts Center)

1830 - 1900 SOCIAL (Cash Bar)

1900 - BANQUET

Thursday, 23 October 1997

0830 - 0930 GENERAL SESSION III

Chair: Robert Launer, U.S. Army Research Office

BAYESIAN STATISTICS: ATTITUDES AND METHODS
Donald Berry, Duke University

0930 - 1000 Break

1000 - 1200 CONTRIBUTED SESSION III (in parallel with Contributed Session IV)

Chair: David Webb, U.S. Army Research Laboratory

A CALIBRATION STUDY OF THE MOBILE ARMY CAMOUFLAGE EVALUATION (MACE) SYSTEM USING HUMAN CAMOUFLAGE EFFECTIVENESS JUDGMENTS
Kragg P. Kysor, Jock O. Grynovicki, Michael G. Golden
U.S. Army Research Laboratory

DETECTING AND IDENTIFYING INTERACTION IN A TWO-WAY LAYOUT WITH ONE OBSERVATION PER CELL
Robert L. Launer, U.S. Army Research Office, George Washington University

THE UPSIDE-DOWN FUNNEL
Barney Bissinger and John Boyarski, Naval Inventory Control Point

EXPLORATION OF SATELLITE IMAGES IN THE DYNAMICALLY LINKED ARCVIEW/XGOBI/XPLORE ENVIRONMENT
Juergen Symanzik, George Mason University
Dianne Cook, Iowa State University
Sigbert Klinke, Humboldt-Universitaet zu Berlin
N. Lewin, Iowa State University

1000 - 1200 CONTRIBUTED SESSION IV (in parallel with Contributed Session III)

Chair: Jay Conover, Texas Tech

ADVENTURES IN DISCRIMINANT ANALYSIS
David Marchette and Jeff Solka, Naval Surface Warfare Center

COMBINATORIAL TESTS FOR VALIDATION OF CLUSTER ANALYSES
Bernie Harris, University of Wisconsin-Madison

BEST SUBSET SELECTION USING A TWO STAGE GENETIC ALGORITHM
Bradley C. Wallet, Naval Surface Warfare Center

VISUAL EXPLORATORY OF SPATIAL DATA WITH MANET
Adalbert F. X. Wilhelm, George Mason University

Thursday, 23 October 1997 (Continued)

1200 - 1330 Lunch

1330 - 1430 GENERAL SESSION IV

Chair: Bernie Harris, University of Wisconsin-Madison

NONPARAMETRIC DISCRIMINANT ANALYSIS IN HIGH DIMENSIONS VIA GENERALIZED INTERPOINT DISTANCES
Carey E. Priebe, Johns Hopkins University

1430 - 1445 Break

1445 - 1645 CONTRIBUTED SESSION V (in parallel with Contributed Session VI)

Chair: LTC Robert Hammell, U.S. Army Research Laboratory

SOFT COMPUTING, MODELING AND FUNCTION APPROXIMATION
Thomas Sudkamp, Wright State University

FUZZY CONTROL OF ROBOT COMMUNICATIONS
Aivars Celmiņš, U.S. Army Research Laboratory

APPLICATIONS OF FUZZY LOGIC TO STATISTICAL ANALYSIS
Dan Ralescu, University of Cincinnati

REAL-TIME ANIMATION OF PARTICLE BEHAVIORS
Jim X. Chen and Edward J. Wegman, George Mason University

1445 - 1645 CONTRIBUTED SESSION VI (in parallel with Contributed Session V)

Chair: Ann Brodeen, U.S. Army Research Laboratory

WAS EINSTEIN RIGHT?
Henry C. Alberts, Defense Systems Management College, Fort Belvoir

PREDICTIVE STATISTICAL PROCESS CONTROL FOR THE MILITARY
LTC David H. Olwell, Department of Mathematical Sciences, U.S. Military Academy

METHODOLOGY FOR ASSESSING THE CONTRIBUTION OF DIGITIZATION IN THE BATTLEFIELD
Michael Golden, Jock O. Grynovicki, Kragg Kysor, Madeline Swann
U.S. Army Research Laboratory

REGION OF INTEREST IDENTIFICATION IN AERIAL IMAGERY
Jeff Solka, Naval Surface Warfare Center

Friday, 24 October 1997

0830 - 0915 GENERAL SESSION V

Chair: Ed Wegman, George Mason University

NOVEL VISUAL REPRESENTATIONS OF MOVEMENT
J. David Cooke, University of Western Ontario

0915 - 0930 Break

0930 - 1130 CONTRIBUTED SESSION VII

Chair: Carl Russell, Joint National Test Facility

USING PEARSON AND SPEARMAN STATISTICS TO LOOK FOR A DEPENDENCE RELATIONSHIP BETWEEN TWO VARIABLES
Lorrie L. Hoffman, University of Central Florida
Dan Corson, Florida A&M University

THE JOINT DISTRIBUTION OF THE MEAN AND AN EXTREMUM OF A NORMAL SAMPLE, WITH APPLICATIONS TO QUALITY CONTROL
Mark Vangel, National Institute of Standards and Technology

APPROXIMATE QUANTILES FOR THE MULTIVARIATE STUDENTIZED RANGE IN THE CASE OF THREE UNEQUAL GROUPS
Otto Schwalb and James R. Thompson, Rice University

AN IMAGE-BASED GRAND TOUR APPLIED TO MINE-FIELD DETECTION
Wendy Poston, Naval Surface Warfare Center
Ed Wegman, George Mason University

1130 - 1145 Break

1145 - 1245 GENERAL SESSION VI

Chair: Barry A. Bodt, U.S. Army Research Laboratory

RECENT DEVELOPMENTS IN ROBUST DESIGN
Eric Lagergren, National Institute for Standards and Technology

1245 ADJOURN

A CALIBRATION STUDY OF THE MOBILE ARMY CAMOUFLAGE EVALUATION (MACE) SYSTEM USING HUMAN CAMOUFLAGE EFFECTIVENESS JUDGMENTS

Kragg P. Kysor, Jock O. Grynovicki, and Michael G. Golden
U.S. Army Research Laboratory - Human Research and Engineering Directorate
Aberdeen Proving Ground, MD 21005

ABSTRACT

The purpose of this report is to provide the results of a calibration study performed by the the Human Research and Engineering Directorate (HRED) in accordance with a Technical Program Annex (TPA) agreement between the Natick Research, Development, and Engineering Center (NRDEC) and HRED. This report describes the objectives, methodological approach, apparatus, experimental design, data collection procedures, and the results of the study.

INTRODUCTION

The human perception of camouflaged targets is influenced by a number of variables. Examples include the type of camouflage pattern, the type of background, the density of the background, as well as the type and direction of illumination. Typically, evaluations of prototype camouflage patterns are performed in field trials through the use of numerous groups of observers who rate the effectiveness of various patterns against different backgrounds. Large field studies are difficult to conduct as they involve the problem of finding suitable test sites, the performance of many test observations, unpredictable weather, and variations in foliage resulting from the change of seasons. Hence, the Natick Research, Development and Engineering Center (NRDEC) developed the Mobile Army Camouflage Evaluation (MACE) system as a cost effective means of automating the process of evaluating camouflage effectiveness. The benefit of this research for the Department of Defense is to reduce the time and cost of field testing existing and proposed camouflage patterns by providing a portable computerized system that simulates human perception and judgment of camouflage effectiveness in field settings.

The MACE system has the capability of providing the U.S. Army with a field-portable device to objectively evaluate candidate camouflage ensembles. The system assesses camouflage effectiveness by using optically-filtered and digitized video images that are subsequently processed by computer algorithms in an attempt to model human visual perception as it relates to camouflage phenomena. To calibrate the MACE device, the Human Research and Engineering Directorate (HRED) of the Army Research Laboratory (ARL) assisted NRDEC personnel in conducting a calibration study This report presents the MACE measurements correlated with the soldiers' ratings of camouflage effectiveness in order to calibrate MACE measures as predictors of soldier performance.

Overview

The MACE system consists of hardware and software for analyzing digitized video images of camouflaged targets and their backgrounds. The purpose of the MACE system is not to search for targets, but to measure the camouflage effectiveness of a target against its background simulating human perception. The system collects 18 basic optical measures (Kilian, 1992) of a target and its background. Each of these optical measures consists of three levels of measurement related to the CIELAB color description coordinates as standardized by the Commission Internationale de L'Eclairage of 1931. These levels are designated by "L*" (lightness level), "A*" (red-green scale value), and "B*" (yellow-blue scale value). Consequently, the total number of measures or variables acquired for a single MACE system observation is 54 (i.e., 18 basic optical measures x 3 levels L, A, & B). However, presently, the 54 variables are treated as having equal importance in predicting human performance. It is likely that there are differences in the predictive values of these various variables. Consequently, empirical data had to be obtained to determine these differential weights.

Objective

The objective of the study was to calibrate the MACE system by determining the relative importance of each of the 54 MACE variables in predicting human subjective ratings of camouflage effectiveness. Specifically, weighting coefficients for MACE system variables will be based upon soldier ratings of the effectiveness of five different camouflage patterns correlated with the MACE system variables of the target ensemble patterns and their associated backgrounds.

Additionally, the study sought to quantify the effect of the direction of illumination on the human observer ratings of camouflage effectiveness of the candidate ensembles. The research hypothesis was that there would be differences in the human judgments of the camouflage effectiveness of the various ensembles when the camouflage targets and background are front lit versus back lit. The statistical null hypothesis was that there would be no difference in human judgments when the camouflage targets and background are front lit versus back lit.

METHOD

Apparatus

MACE System--The MACE system (Kilian, 1992) studied used a video camera, a series of ten optical filters, and a computer to digitally store an image containing the camouflaged target and its background. Each of the ten filters sampled a different wavelength from the electromagnetic spectrum (380 nm to 740 nm). For each image stored, an outline of the camouflaged target portion of the image was made and then the digitized pixel information within the target outline was stored separately from the background information portion. Next, 54 variables were computed that were thought to differentiate the target from the background. These variables were based on features such as light reflectance, spectral wavelength, and texture. During the study, the MACE equipment was operated and maintained by NRDEC personnel.

Targets--The calibration study compared five candidate camouflage patterns. The patterns were woodland, olive drab, desert, solid gray, and black. Though one would expect the woodland pattern to be most effective against a wooded background, the traditional olive drab continues to show its concealment value under certain conditions (Bensel, Ramsley, & Bushnell, 1977). The solid gray ensemble has been proposed for use in built-up and urban areas but may also be effective in wooded situations. The desert pattern may also be effective in the woods where there is sparse foliage and bright sunlight with a dried grassy background. The black pattern, intended for night use, may also be effective in dense and shaded wooded areas.

Target Backgrounds--The backgrounds of the target ensembles consisted of homogeneous grass, trees, and their associated foliage. The tree trunks and branches were brownish-gray while the foliage consisted of clusters of mostly green leaves.

Observation Stations--An observation station, consisting of 10 individual observer cubicles, was provided for the test participants to view the pairs of camouflage ensembles. The individual observation cubicles were used (as opposed to a group observation and judgment situation) to obtain independent test participant judgments of the camouflage effectiveness of the ensembles. The test participants sat in the cubicles which blocked their vision down range. When signaled, the test participants stood to observe the camouflage target ensemble pair for 10 seconds and then were seated.

Test Site Layout --The overall test site layout is presented in Figure 1. The Figure shows a distance of 30 meters from the MACE apparatus and the test observers to the target camouflage ensembles. The distance between the observer and the target was considered critical to the calibration process. It is a generally accepted principle that the details of an image of an object are less discriminated as the distance between the observer and the object are increased. Likewise, the MACE system resolution decreases as a function of distance. Preliminary investigations at Natick showed a target filled one-half of the vertical extent of the MACE computer monitor height when the MACE to target distance was approximately 13 meters. It did not seem reasonable to expect the MACE system to provide optimal data when the target height was less than 20% of the MACE display height. The size of the target area that is represented by one MACE camera pixel is approximately one centimeter at 20 meters and two centimeters at 40 meters. However, at a distance of 40 meters and beyond, the target height would be less than 20% of the MACE display height and so small that only large areas of contrast would be discriminated. Thus, it was decided that a distance of 30 meters be used to provide a level of sensitivity at which the MACE system could adequately discriminate the detail that is found in woodland and desert camouflage patterns.

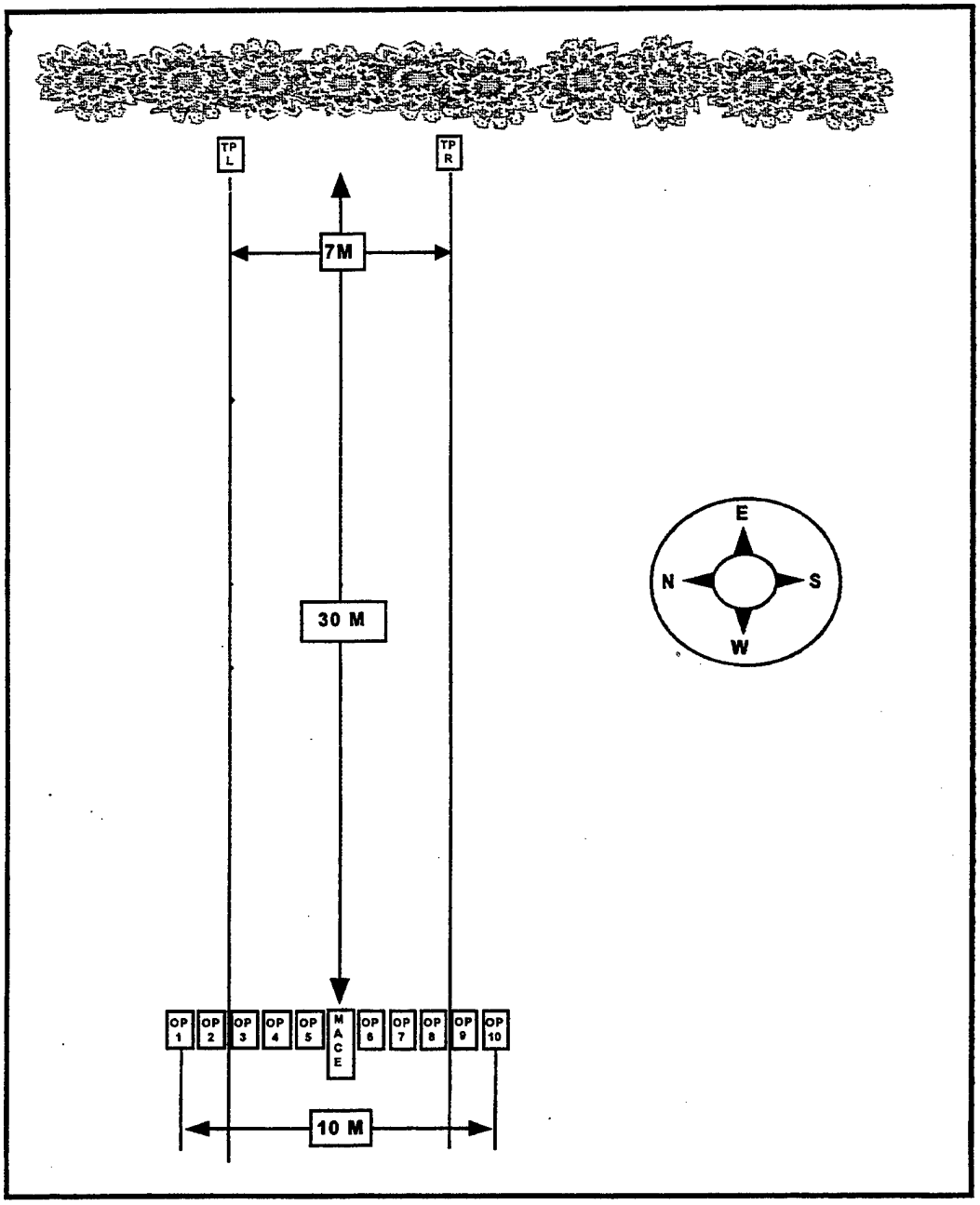


Figure 1. Diagram of test area (overhead view) showing relationship of the left and right camouflage patterned target ensemble positions (TPs) to the MACE system and the ten observation positions (OPs) backgrounds. The view of the target ensembles and their backgrounds, as seen by the test observers and the MACE system video camera, is shown in Figure 2.

Left and right target ensembles, spaced 10 meters apart, were presented to correspond with two tree foliage backgrounds that were homogeneous. Front lit illumination occurred when the sun position at the test site (see Figure 1) was in the west and the resulting shadows, from the observer's viewpoint, were behind the target patterns and the corresponding backgrounds. Back lit illumination occurred when the sun was in the east and the

resulting shadows were in front of the target patterns and backgrounds. The view of the target ensembles and their backgrounds, as seen by the test observers and the MACE system video camera, is shown in Figure 2.

Figure 2. The diagram shows a simulated MACE system computer video image display of two target camouflage ensembles with similar foliage backgrounds. The white outline around each target shows the background area of interest (AOI) that was used for acquiring the data to compute the MACE system variables' values. A reflectance board with eight light levels was used to standardize the MACE video camera input.

Test Participants

Ten male enlisted U.S. Army soldiers (ages ranging from 19 to 35 years) participated as judges of the camouflage effectiveness offered by the five camouflage patterns. Each candidate observer was tested for visual acuity and color deficiency. Only candidates whose visual acuity was measured at 20/20 or better (corrected or uncorrected) were selected to participate in the study.

Test Participant Procedure

Each test participant was given a response sheet at the beginning of each trial. The test participant placed a check mark in one box of a sequence of seven boxes corresponding to a 7-point rating scale (1=least blending, 4=moderate blending, 7=most blending) that indicated the degree that he perceived that the left-sided target camouflage ensemble pattern blended with the background. Similarly, he rated the right-sided target ensemble pattern.

MACE Data Acquisition Procedure

The MACE system data was obtained simultaneously with the collection of the test participant data. A reflectance board was located at the base of the target camouflage ensembles to calibrate the MACE apparatus for standardized video data input. If necessary, MACE filters were changed to adjust for the variation in lighting conditions at the time of the data acquisition.

Experimental Methodology

Experimental Conditions--The experimental conditions for the study were the type of camouflage (Woodland, Olive Drab, Desert, Solid Gray, and Black), sun illumination condition (Front lit and Back lit), and target-background position (Left and Right). The same group of soldier test participants was used under all the experimental treatment conditions which resulted in a repeated measures 2 x 5 x 2 design (Direction of Illumination x Camouflage Pattern x Left-Right Target-background Position). These experimental conditions for obtaining the soldier camouflage effectiveness judgments and MACE measurements are presented in Table 1.

Table 1

Experimental Conditions for Obtaining the Soldier Camouflage Effectiveness Judgments and MACE System Measurements

Direction of Illumination										
Front-Lit						Back-Lit				
Camouflage Pattern						Camouflage Pattern				
*W R/L	O R/L	D R/L	G R/L	B R/L		W R/L	O R/L	D R/L	G R/L	B R/L

*Camouflage Codes: W=Woodland O=Olive Drab D=Desert G=Gray B=Black
Position Codes: R=Right L=Left

Dependent Variable--The dependent variable was the human camouflage effectiveness response obtained from the soldier participant observer. This measure was a rating, on a 7-point scale (1=least blending, 4=moderate blending, 7=most blending), of the camouflage effectiveness of each ensemble pattern against the background.

A paired comparison method was used to obtain the human response measure for this study. The paired comparison method is "a procedure in which objects are compared with each other in pairs, each with each, till all combinations are given" (English & English, 1958). This method was chosen because it allowed the test participant to concentrate on the immediate perceptual task and not rely on memory as is required by some other methods.

The paired comparisons of five test ensembles resulted in ten combinations of pairs. However, to control for variations in the local background, each pair of camouflage ensembles was presented twice; that is, once with a member of the paired combination viewed on the left position and once with the member viewed on the right position (see Figures 1 & 2). Because each pair of ensembles was presented twice (left and right positions) the number of paired combinations was doubled to 20 presentations. The presentation of these 20 pairs of ensembles was randomized within each session to control for biased responses that may result from learned expectancies resulting from systematic patterns of ensemble presentations. The time between left and right presentations was approximately 100 minutes. The result of this paired comparison procedure allowed the computation of group average ensemble pattern ratings of camouflage effectiveness from the ten test participants for each trial of the ten ensemble pair presentations.

Predictor Variables--The predictor variables are the MACE basic system variables related to light reflectance, spectral wavelength, and texture. Each of the basic variables resulted in three measures at three levels of analysis related to lightness (black-white scale(L)), and color; i.e., a red-green scale value(A) and a yellow-blue scale value(B). A list of the MACE system variables is presented in Table 2. The Bhattacharyya distance measure is described by Fukunaga (1990). The texture variables are based on the reports by Haralick, Shanmugam, and Dinstein (1973) and Connors and Harlow (1980).

Table 2

List of Mobile Army Camouflage Evaluation (MACE) System Variables

Var. No.	Stem Abbrev.	Name of Basic Variable	L (Prefix)	A	B
		Lightness and Color Measures		(Prefix + Stem)	
1.	MEA	Mean (pixels)	LMEA	AMEA	BMEA
2.	VAR	Standard Deviation	LVAR	AVAR	BVAR
3.	SKE	Skewness	LSKE	ASKE	BSKE
4.	KUR	Kurtosis	LKUR	AKUR	BKUR
5.	MIN	Minimum Value	LMIN	AMIN	BMIN
6.	MAX	Maximum Value	LMAX	AMAX	BMAX
7.	BAT	Bhattacharyya Distance	LBAT	ABAT	BBAT
		Texture Measures			
8.	A2M	Angular 2nd Moment	LA2M	AA2M	BA2M
9.	ENM	Entropy	LENM	AENM	BENM
10.	CSM	Cluster Shade	LCSM	ACSM	BCSM
11.	CPM	Cluster Prominence	LCPM	ACPM	BCPM
12.	LCM	Linear Correlation	LLCM	ALCM	BLCM
13.	I1M	Info Measure of Correlation I	LI1M	AI1M	BI1M
14.	I2M	Info Measure of Correlation II	LI2M	AI2M	BI2M
15.	SGM	Angular 2nd Moment of SGLDM	LSGM	ASGM	BSGM
16.	ESM	Entropy of SGLDM	LESM	AESM	BESM
17.	S2M	2nd Moment of SGLDM	LS2M	AS2M	BS2M
18.	ISM	Inverse 2nd Moment of SGLDM	LISM	AISM	BISM

Note. The abbreviation of a MACE system variable consists of a prefix, a stem, and a suffix. The prefix is the letter L, A, or B indicating lightness (L), red-green scale (A), or yellow-blue scale (B) variable. The stem is a three-character abbreviation of the basic MACE system name (see Table 2). The suffix designation, used later in the report, is a one-character number indicating a quadratic (2) or cubic (3) transformation of the MACE system variable.

Data Analyses

Ensemble Camouflage Effectiveness Ratings. The 10 test observers viewed each of the camouflage ensemble pairs 20 times during the course of the 10 sessions. Each observer rated the camouflage ensemble, on a 7-point scale, for the degree that the ensemble blended with the background. A "1" on the scale indicated low camouflage effectiveness, a "4" indicated moderate camouflage effectiveness, and a "7" indicated high camouflage effectiveness.

MACE Predictor Weights and Ensemble Camouflage Effectiveness Ratings. The MACE system variables were considered as independent or "predictor" variables and the soldier ensemble effectiveness rating was considered as the dependent or "outcome" variable. The variables that contributed most to the explanatory variance were studied to try to understand the underlying psychophysical processes involved. This subset of variables was then considered as a smaller set of predictor variables to simplify the complexity of the multivariate interrelationships. The determination of the relative weighting coefficients for the MACE system factors was accomplished by multiple regression analyses. It is acknowledged that the camouflage effectiveness ratings are ordinal level measurements as opposed to the more desired interval level measurements for multiple regression analyses. However, at this early stage of machine perception research, the multiple regression method was felt to be sufficiently robust to allow the identification of the relative importance of the MACE system variables.

In an attempt to validate the multiple regression prediction equations within the parameters of the study, a portion of the data was randomly deleted and the remaining data used to develop the equations. The Pearson correlation coefficient between the observed and predicted responses was obtained to ascertain the reliability of the weighting coefficients.

Experimental Treatments. The mean ensemble 7-point camouflage effectiveness ratings for the individual ensemble patterns were computed for the front and back lit illumination conditions and the left and right target-background positions. The statistical significance of differences between the means was detemined by analyses of variances. If there were no significant differences, then the predictor weights would be assumed to be valid for the experimental treatment conditions used. However, if there were significant differences, then further analyses and studies focusing on the specific treatment conditions would be indicated.

RESULTS

Effect of Camouflage Pattern, Light Direction, and Target-Background Position

The raw data for the 54 MACE system variables and soldier observer ratings of camouflage effectiveness were obtained from the total of 345 trials. The mean ratings for each camouflage pattern are presented, graphically, in Figure 4. The resulting rank order for the five camouflage patterns against a woodland background was (1) woodland, (2) olive drab, (3) black, (4) gray, and (5) desert. The mean ratings for each camouflage pattern were further analyzed by the effect of light direction (front and back) and target-background position (left and right). These data are presented graphically in Figure 5. It can be seen that there is virtually no effect of light direction or target-background position on the soldier response ratings.

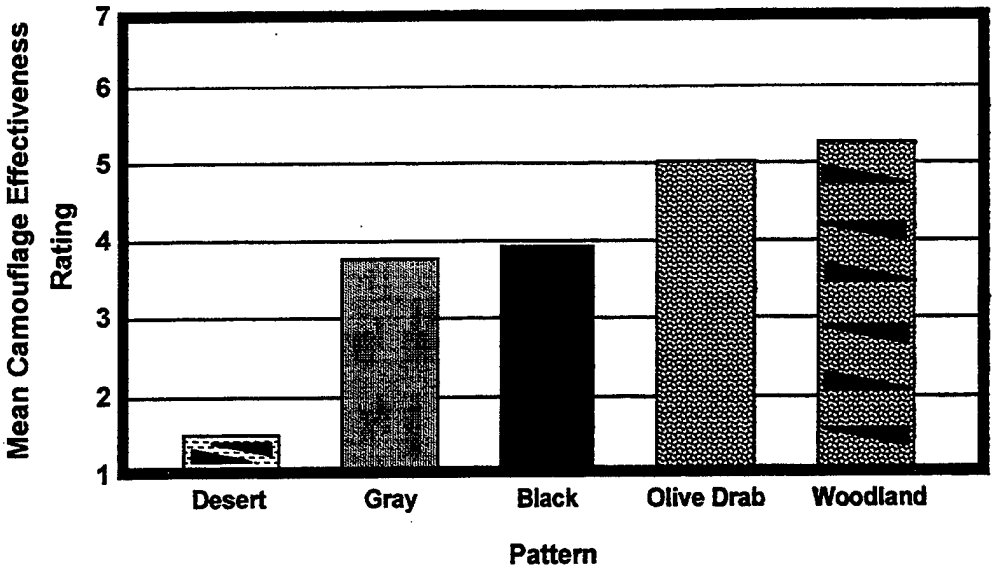


Figure 4. Soldier rating of camouflage effectiveness. (1=least blending, 4=moderate blending, 7=most blending).

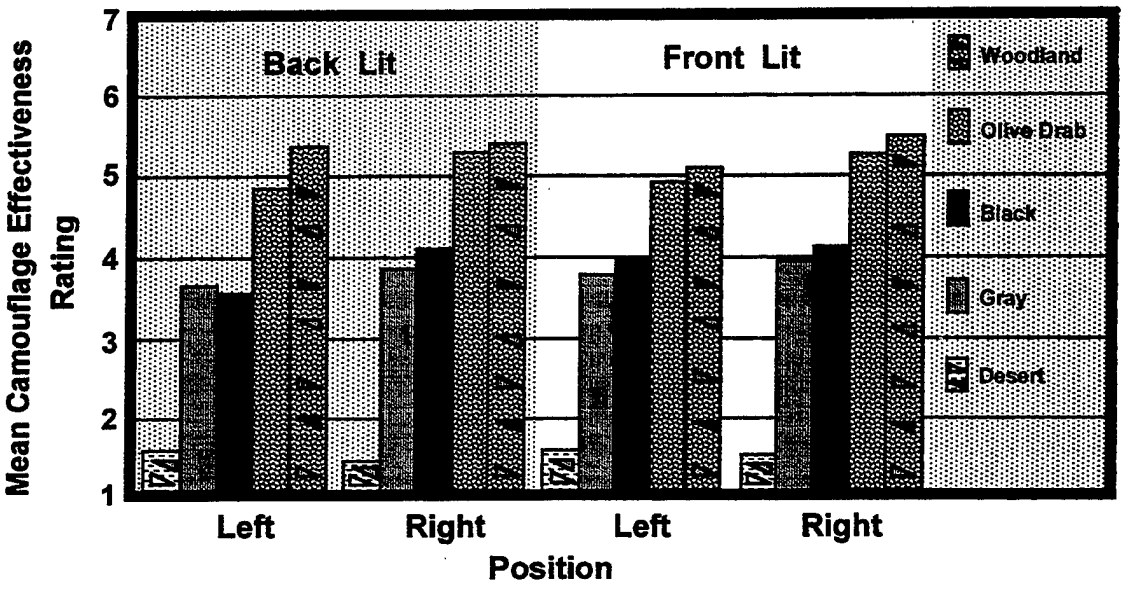


Figure 5. The effect of the position of the target (left or right side) by direction of lighting (front lit vs. back lit) and by camouflage ensemble pattern on the mean camouflage effectiveness rating.

Relative Importance of the MACE System Variables

In order to determine the relative importance of the MACE system variables in predicting the soldier rating responses, the stepwise technique of the multiple regression analysis method was used (SPSS Release 4.0 for Sun 4). Prior examination of the raw data indicated that some of the relationships between the MACE system variables and the human response variable (rating) were nonlinear; therefore, quadratic and cubic transformations of the raw data were also used in the subsequent multiple regression analyses. Thus, the 54 L, A, and B variables (Table 2) and their quadratic and cubic transformations resulted in a total of 162 MACE system candidate predictor variables. When the 162 candidate predictor variables were correlated with the soldier ratings of camouflage effectiveness, 120 of these candidate variables were found to be statistically significant. Stepwise multiple regression analyses considering the entire pool of 162 candidate variables resulted in an optimum subset of 16 predictor variables with a multiple R of 0.95 ($F(16)=189.7$, $p<.01$). The beta coefficients show that variables relating to the overall lightness difference between the target and background contributed most of the explanatory variance. The next most important variables were those relating to the yellow-blue color scale.

The stepwise progression of the multiple R is presented in Figure 6. The pattern of the growth curve of the multiple R as variables are selected by the stepwise method shows an initial correlation of .827, a rapid increase as the next few variables are considered, and finally, a leveling of the curve as the maximum number of statistically significant variables is achieved (a multiple R of .95 for the 16 specific predictor variables).

Obtaining an Efficient Prediction Model

To minimize the effect of multicollinearity that results from the use of too many predictor variables, a parsimonious subset consisting of the first 6 variables of the 16 specific predictor variables was chosen as a candidate model for prediction of human performance from the MACE system. These 6 predictor variables; i.e., LMEA2, LMEA3, BMEA, BVAR, BMEA3, and LMAX produced a multiple R_6 of .93 ($F(6)=332.6$, $p<.01$) which is almost as high as the multiple R_{16} of .95 obtained when the 16 variables are used. The weighting coefficients for the subset of 6 variables is presented in Table 3.

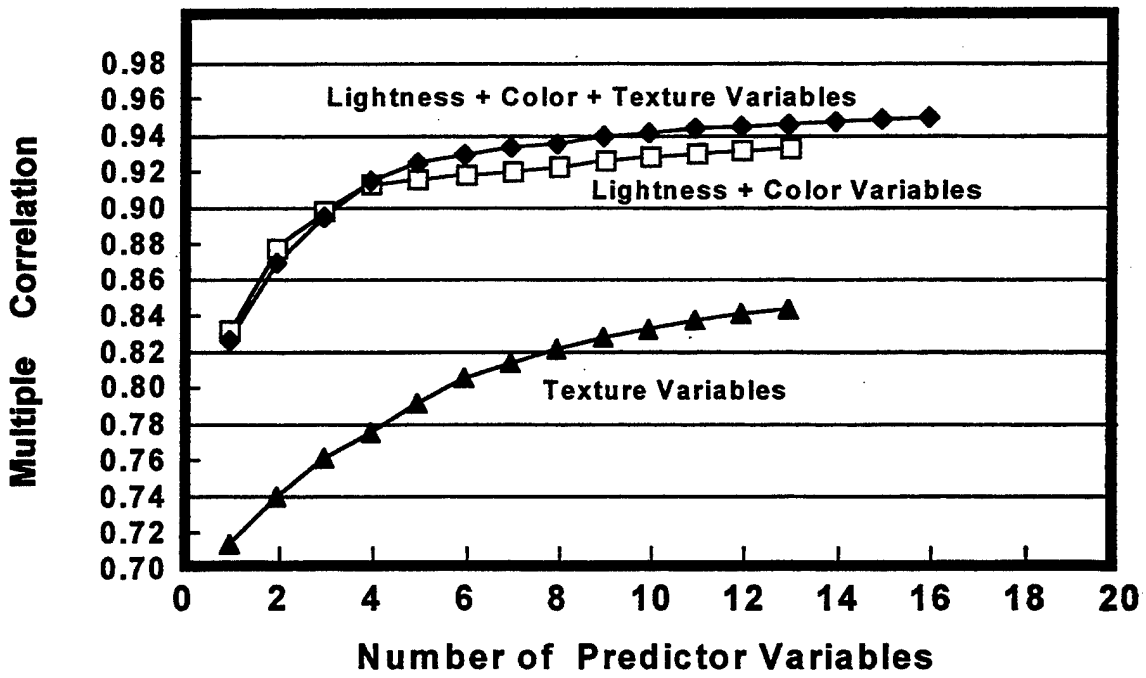


Figure 6. The effect of the number of predictor variables and type of variable (lightness, color, or texture) on the size of the multiple correlation.

To test the reliability of this parsimonious six-variable predictor model, the original raw data set of 345 observations was divided into two parts. The first part consisted of 305 observations that formed a prediction data subset (Group 1) that was used to obtain a new set of predictor weighting coefficients for the six variables. The second part consisted of the remaining 40 observations (Group 2) that were previously randomly selected from the 345 observations. The new set of weighting coefficients obtained from Group 1 was then applied to the same six variables of the 40 observations of Group 2. The prediction equations were then solved for the Group 2 observations which yielded predicted camouflage ensemble ratings for this verification group that were compared with the actual ratings of this group. The Pearson correlation coefficient obtained from the comparison of the predicted and actual ratings was an *r* of 0.90. The correlation and regression between the predicted and actual soldier camouflage ratings for this parsimonious six-variable predictor model is shown in Figure 7. The close fit suggests that this parsimonious version is an efficient and useful model for predicting human ratings of camouflage effectiveness.

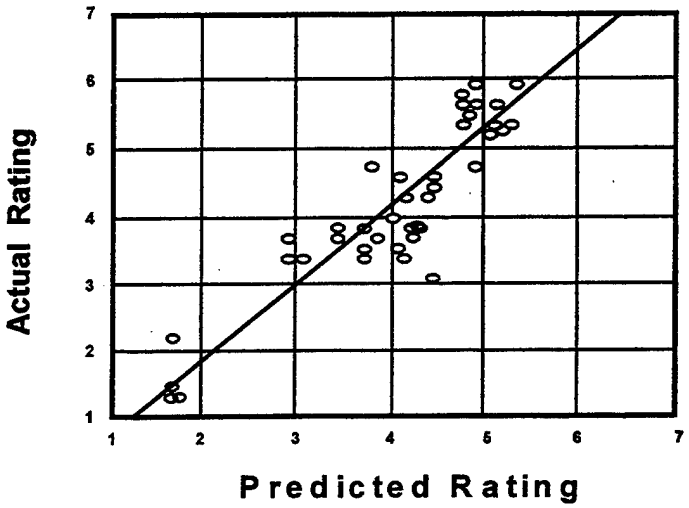


Figure 7. Plot shows the correlation between the predicted and the actual soldier camouflage effectiveness ratings. The model is based on six MACE system predictor variables (LMEA2, LMEA3, BMEA, BVAR, BMEA3,LMAX). Forty cases plotted. Correlation = .90 R Squared =.81 S.E. of Est.=.54 2-tailed Sig.=.000 Intercept (S.E.)= -.15 (.35) Slope (S.E.)= 1.07 (.08)

TABLE 3

Weighting Coefficients of Six Parsimonious Predictor Variables to Obtain an Efficient Model.

Step	MultR	Rsq	Variable	Type	B	Beta	T	Sig T
1	.836	.698	LMEA2	L	- .003120	-3.479084	-19.7	.000
2	.879	.773	LMEA3	L	3.18050E-05	2.422736	13.6	.000
3	.905	.818	BMEA	B	.081652	.717738	14.8	.000
4	.919	.845	BVAR	B	- .027090	- .164391	- 5.7	.000
5	.928	.862	BMEA3	B	- 2.55107E-05	- .259992	- 6.6	.000
6	.933	.870	LMAX	L	- .006124	- .134595	- 4.3	.000
			(Constant)		4.625755		67.9	.000

Note. MultR = Multiple correlation coefficient; Rsq = the Multiple correlation coefficient squared (explained variance); Type: (L) lightness scale, (A) red-green scale, (B) yellow-blue scale; B = the raw data weighting coefficient (slope); Beta = the standardized weighting coefficient; T = slope / standard error of the slope; Sig T = the statistical significance of T.

The Relative Contribution of Lightness, Color, and Texture Information to the Predictive Models

While the parsimonious six-variable predictor model appeared useful, it was noted that most of the higher-weighted variables were related to lightness and the yellow-blue color scale whereas the lesser-weighted variables tended to be measures of texture. Consequently, two additional stepwise multiple regression analyses were performed to assess the relative contribution of the basic lightness and color variables versus the texture variables. The results of these analyses are also shown in Figure 6. It can be seen that the stepwise multiple regressions based on the lightness and color variables alone produce virtually the same predictive information as when the lightness, color, and the texture variables are used together. In contrast, the multiple regressions based on texture variables alone, are markedly below those based exclusively on the lightness and color variables. From this finding, it was determined that a model based solely upon the lightness and color variables could serve as a useful model for predicting human ratings of camouflage effectiveness.

To further test the effectiveness of a model based on lightness and color variables without the effect of texture, a second model was produced. The variables chosen for this model were the first six variables obtained from the multiple regression analysis that was done considering only lightness and color related variables. These six variables were LMEA, BMEA, BBAT, BVAR, BBAT2, and ABAT2. The weighting coefficients for this lightness and color model are shown in Table 4. The multiple *R* obtained from these six variables was .92. The reliability of the coefficients obtained from these six variables was an *r* of .86. Thus, further evidence was found to suggest that the parsimonious six-variable predictor model which, also, contains only lightness and color variables (see Table 3) is an effective and useful model for predicting human ratings of camouflage effectiveness.

TABLE 4

Weighting Coefficients of Six Predictor Variables to Obtain a Model Based on only Light and Color Variables (versus Texture Variables)

Step	MultR	Rsq	Variable	Type	B	Beta	T	Sig T
1	.843	.711	LMEA	L	- .076475	-1.226763	-26.0	.000
2	.891	.794	BMEA	B	.072358	.636039	13.3	.000
3	.909	.827	BBAT	B	- 2.475205	- .370344	- 5.2	.000
4	.920	.846	BVAR	B	- .030880	- .187390	- 6.3	.000
5	.922	.850	BBAT2	B	1.283458	.176110	2.5	.012
6	.924	.854	ABAT2	A	.564369	.080290	2.8	.005
				(Constant)	5.560999		72.1	.000

Note. MultR = Multiple correlation coefficient; Rsq = the Multiple correlation coefficient squared (explained variance); Type: (L) lightness scale, (A) red-green scale, (B) yellow-blue scale; B = the raw data weighting coefficient (slope); Beta = the standardized weighting coefficient; T = slope / standard error of the slope; Sig T = the statistical significance of T.

DISCUSSION

The results of the various statistical analyses described herein indicate that, of the L (lightness scale), A (red-green scale), and B (yellow-blue scale) variables, the most predictive were those involved with the lightness scale, followed by those concerned with the yellow-blue color scale. The importance of the lightness scale variables suggests that a target may be camouflaged to a large extent simply by the closeness of its reflected gray-scale light level pattern to the overall background gray-scale pattern. As for the role of the yellow-blue color scale variables, given the woodland background used in this study (which exhibited more of the colors in the yellow-blue scale than the red-green scale) it is reasonable to find that the human rating of high camouflage effectiveness would be associated with camouflage ensemble colors patterns most similar to the colors in the woodland background.

It was interesting, however, to discover that the texture variables were less highly weighted than were the variables involved with lightness and color. This finding leads one to consider how the distance factor may play a role. Generally, the ability of a viewer to discern the details of an object are related to the distance between the viewer and the object (Gibson, 1950). The viewing distance recommended by the developers of the MACE system for the apparatus was 20-100 meters (Goodell & Kilian, 1992). However, to provide some camouflage effect, we felt that 20 meters would be too close; i.e., the target object would look artificial and be too easy for the soldier raters to discern. On the other hand, at the longer distances, the image of the target pattern on the MACE computer display would involve so few pixels, the opportunity for the MACE system computer algorithms to discriminate among the patterns would be minimal. Consequently, the distance of 30 meters was chosen as being just far enough away to provide some camouflage effect of the target patterns, yet would still provide the MACE system sufficient target pattern pixels to allow the computer algorithms to analyze textural differences. While the 30 meter distance selected appeared optimum it may still not have allowed the textural variables to contribute sufficient information to discriminate among the patterns used.

In conclusion, the objectives of this study were to determine whether or not the MACE system variables could discriminate among a range of camouflage patterns and correlate with human judgments of camouflage effectiveness. The results of the investigation provided a combination of 16 weighted MACE system variables that correlated highly with the soldiers' pattern ratings. Further analyses yielded an even smaller number of MACE system variables (the six-variable parsimonious predictor model), which involved only lightness and color variables, and provided virtually the same predictive information as the 16 variable model. It was also suggested that the MACE system could be further enhanced to exploit camouflage texture information at close ranges which might further increase the system's performance in predicting human perception of camouflage effectiveness and perhaps lead to the development of improved camouflage textural patterns. These possibilities deserve further investigation. Overall, based on the results of this study, the MACE system concept, as a cost-effective means of automating the process of evaluating camouflage effectiveness, appears promising.

REFERENCES

Bensel, C., Ramsley, A.O., and Bushnell, W.B. (1977). A comparative study of matrix measures for maximum likelihood texture classification. *IEEE Transactions of Systems, Man, and Cybernetics, 21*(1), Jan./Feb.

Connors, R. and Harlow, C. (1980). A theoretical comparison of texture algorithms. *IEEE Trans. Patt. Analysis & Mach. Intell., PAMI-2,* 204.

English, H.B. and English, A.C. (1958). *A comprehensive dictionary of psychological and psychoanalytical terms.* New York: David McKay Co.

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition,* 2nd Ed., Boston: Academic Press.

Gibson, J.J. (1950). *The perception of the visual world.* New York : Houghton Mifflin Co.

Goodell, G.F. and Kilian, J.C. (1992). *User's Manual for Mobile army camouflage evaluation (MACE) system (v.2.0).* Report No. COI-SR-1031-43 Prepared for U.S. Army Natick RD & E Center, MA, 28 Feb 92.

Haralick, R., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Trans Sys Man & Cyber., SMC-3,* 610.

Kilian, J.C. (1992). *Mobile army camouflage evaluation (MACE) system.* Final Report No. COI-SR-1031041.2 Prepared for U.S. Army Natick RD & E Center, MA, May.

Detecting and Identifying Interaction in a Two-way Layout With One Observation per Cell

Robert L. Launer *
George Washington University, and
Army Research Office

ABSTRACT

The analysis of variance procedure produces no estimate of interaction when there is insufficient repetition within cells to estimate error. An example of this is the two-way layout with one observation per cell. Johnson and Graybill presented a method for that situation which allows the identification of those cells which may contain interaction. The identified cells are excluded from a subsequent analysis of variance.

In this paper, an alternate method is presented for the identification of those cells in a two-way layout with one observation per cell with interaction, which also produces an estimate of the interaction or overall deviation from additivity. The method produces n (= at) "t-statistics", one for each cell, for determining the significance of the estimates. Although not necessary, it is prudent to assume that the number of cells which contain interaction is small.

INTRODUCTION.

The two-way analysis of variance for data which may contain interaction has produced a small body of interesting literature. Tukey[1] was the first to propose a method for determining whether interaction exists in this situation. Tukey's method involves the insertion of a single parameter into the model which requires only one degree of freedom for testing for no interaction. Let

$$y_{ij} = \mu + \alpha_i + \tau_j + \lambda\alpha_i\tau_j + \varepsilon_{ij}$$

where $i = 1,2, ..., a; \quad j = 1, 2, ..., t;$

ε_{ij} are NID$(0, \sigma^2)$; and $\sum_{i=1}^{a}\alpha_i = \sum_{i=1}^{t}\tau_j = 0.$

In this model, λ is a "global" interaction coefficient which does assume a certain structure. Several variants of this model have been proposed and studied, while other authors have studied the problem of identification of the specific cells which contain interaction. A review of this literature may be found in Johnson and Graybill[2].

Johnson and Graybill[2,3] produced a satisfactory solution to this problem. After applying Tukey's test to determine whether or not interaction does exist,

---

* The research for this paper was partially completed while the author was at the University of South Carolina, Columbia, South Carolina.

they estimate the interaction in all possible 2x2 contrasts, and declare a cell to contain interaction if it is contained in a large number of those contrasts which produce a significant statistic. This is a significant step, but the total number of 2X2 contrasts is $\binom{a}{2}\binom{t}{2}$. This quantity is always larger than n = at for a > 3, and t > 3, and can be much larger than n for even moderately large a and t. Since each of the n cells is contained in (a - 1)(t - 1) distinct 2x2 contrasts, judgement is required to determine if a particular cell should be declared "significant" or not. The simultaneous inference problem is avoided by the examination of all of the $\binom{a}{2}\binom{t}{2}$ statistics. Furthermore, no follow-up analysis is required since all of the information is provided in one application of their method. Their method does not produce an estimate of the size of the interaction. This is not a strong criticism, but some sort of estimate of the non-additivity would be useful.

In this paper, a method is proposed which makes no assumptions about the structure of the interaction. It produces individual estimates and associated "t-statistics" for interaction in each of the n = at cells of the two way layout. The issue of simultaneous inference is not completely eliminated since some follow-up analysis may be desired if the total number of cells containing interaction is large. That is, the t-statistic in a given cell is computed with the error estimate based on the observations in all of the other cells. If several cells are discovered to contain interaction, it might be desirable to conduct a final test based on the removal of all of these cells.

THE GENERAL CASE

A general solution to the problem is presented here, after which several special cases will be examined. The general notation is that of Graybill[4]. To motivate additional notation, notice the following design matrix. This matrix represents a 3x3 array with one observation per treatment combination in which cells (1,1) and (1,2) include interaction. That is, two observations involve interaction. Notice the 2x2 identity matrix in the upper right hand corner of this matrix, and notice the 5x2 matrix of 0's just below it. The model which would generate this design matrix is:

$$y_{ij} = \begin{cases} y_{ij} = \mu + \alpha_i + \tau_j + \lambda_{ij} + \varepsilon_{ij}, & i,j = 1,1 \text{ or } 1,2 \\ y_{ij} = \mu + \alpha_i + \tau_j + \varepsilon_{ij}, & \text{otherwise} \end{cases}$$

μ	α_1	α_2	α_3	τ_1	τ_2	τ_3	γ_{11}	γ_{12}
1	1	0	0	1	0	0	1	0
1	1	0	0	0	1	0	0	1
1	1	0	0	0	0	1	0	0
1	0	1	0	1	0	0	0	0
1	0	1	0	0	0	1	0	0
1	0	0	1	1	0	0	0	0
1	0	0	1	0	0	1	0	0

More generally, let y_{ij}, i = 1, 2, ..., a and j = 1, 2, ..., t represent n = at

observations of which q involve interaction. (Later, restrictions on the sizes of a, t, and of q relative to a and t which will allow testing will be apparent.) Let Γ represent the set of the cells in which interaction occurs. Then

$$y_{ij} = \begin{cases} y_{ij} = \mu + \alpha_i + \tau_j + \lambda_{ij} + \varepsilon_{ij}, & i,j \in \Gamma \\ y_{ij} = \mu + \alpha_i + \tau_j + \varepsilon_{ij}, & i,j \notin \Gamma \end{cases}$$

It will be convenient to represent these models in matric form as follows. Let y represent the vector of observations and let $y = \begin{bmatrix} y_\gamma \\ y_\beta \end{bmatrix}$, where y_β contains those observations in y which involve no interaction, and y_γ contains those observations in y which involve interaction. Let β' = represent the px1 vector with components $(\mu, \alpha_1, \alpha_2, \ldots \alpha_a, \tau_1, \tau_2, \ldots \tau_t)$, where $p = (a + t + 1)$ and let γ represent the vector of all interaction parameters in the model, where q is the total number of these parameters. Let X_γ be the qxp matrix which corresponds to those observations which involve the interaction parameter, γ, and let X_β be the (n-q)xp matrix corresponding to those observations which do not involve the interaction parameter. Then the design matrix, in partitioned form, appears as follows (where 0 is an (at-q)xq matrix of 0's and I_q is the qxq identity matrix):

$$X = \begin{bmatrix} X_\gamma & I_q \\ X_\beta & 0 \end{bmatrix}$$

The model may be represented as,

$$y = \begin{bmatrix} y_\gamma \\ y_\beta \end{bmatrix} = \begin{bmatrix} X_\gamma & I_q \\ X_\beta & 0 \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \varepsilon$$

where ε represents the vector of errors. A formal manipulation results in the following,

$$X'X = \begin{bmatrix} X_\gamma'X_\gamma + X_\beta'X_\beta & X_\gamma' \\ X_\gamma & I_q \end{bmatrix}$$

and the normal equations are:

$$\begin{bmatrix} X_\gamma'X_\gamma + X_\beta'X_\beta & X_\gamma' \\ X_\gamma & I_q \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} X_\gamma' & X_\beta' \\ I_q & 0 \end{bmatrix} \begin{bmatrix} y_\gamma \\ y_\beta \end{bmatrix}$$

The conditional inverse or c-inverse of $X'X$ denoted by $(X'X)^C$ is:

$$(X'X)^C = \begin{bmatrix} (X_\beta'X_\beta)^C & -(X_\beta'X_\beta)^C X_\gamma' \\ -X_\gamma(X_\beta'X_\beta)^C & I_q + X_\gamma(X_\beta'X_\beta)^C X_\gamma' \end{bmatrix}.$$

Finally, $(X'X)^C X'$ is:

$$(X'X)^{C}X' = \begin{bmatrix} 0 & (X_\beta'X_\beta)^{C}X_\beta' \\ I_q & -X_\gamma(X_\beta'X_\beta)^{C}X_\beta' \end{bmatrix} .$$

Then, the least squares estimator of β and γ are:

$$\hat{\beta} = (X_\beta'X_\beta)^{C}X_\beta' y_\beta$$

and,

$$\hat{\gamma} = y_\gamma - X_\gamma(X_\beta'X_\beta)^{C}X_\beta' y_\beta$$

$$= y_\gamma - X_\gamma\hat{\beta} .$$

Thus the algorithm is a 2-stage operation consisting in obtaining estimators of β using only those observations which contain no interaction, and then insert these into the formula for obtaining the estimate for γ .

The estimate of the error variance is obtained in the usual way. Let θ represent the vector $\begin{bmatrix} \beta \\ \gamma \end{bmatrix}$ and $\hat{\theta}$ represent the corresponding vector of estimates. Then, substituting $\hat{\gamma} = y_\gamma - X_\gamma\hat{\beta}$ into the r.h.s. of the first equality immediately following, the result is:

$$(y-X\hat{\theta})'(y-X\hat{\theta}) =$$

$$\left[\begin{bmatrix} y_\gamma \\ y_\beta \end{bmatrix} - \begin{bmatrix} X_\gamma & I_q \\ X_\beta & 0 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} \right]' \left[\begin{bmatrix} y_\gamma \\ y_\beta \end{bmatrix} - \begin{bmatrix} X_\gamma & I_q \\ X_\beta & 0 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} \right]$$

$$= \begin{bmatrix} 0 \\ y_\beta - X_\beta\hat{\beta} \end{bmatrix}' \begin{bmatrix} 0 \\ y_\beta - X_\beta\hat{\beta} \end{bmatrix} .$$

$$= (y_\beta - X_\beta\hat{\beta})'(y_\beta - X_\beta\hat{\beta})$$

The error variance is, therefore, estimated only with those observations which are not effected by interaction. Note that $(I_{(n-q)} - X_\beta(X_\beta'X_\beta)^{C}X_\beta')$ is symmetric and idempotent. Since the rank of X_β is q less than the rank of X then

$$\hat{\sigma}^2 = \frac{1}{(a-1)(t-1)-q} (y_\beta - X_\beta\hat{\beta})'(y_\beta - X_\beta\hat{\beta})$$

$$= \frac{1}{(a-1)(t-1)-q} y_\beta' (I_{(n-q)} - X_\beta(X_\beta'X_\beta)^{C}X_\beta') y_\beta ,$$

Assuming that the errors are NID(0, $\sigma^2 I$), then $\hat{\sigma}^2$ is distributed as a chi-square random variable with (a-1)(t-1)-q degrees of freedom, and $\hat{\sigma}^2$ is the UMVUE of

σ^2. An examination of the degrees of freedom indicates that in a 2x2 layout, it would be impossible to estimate an interaction coefficient. In a 3-level layout, there are enough degrees of freedom to simultaneaously estimate the interaction coefficients associated with no more than 3 cells. In general, this "saturation" limit is (a-1)(t-1)-1.

It is important to point out that the linear forms $\hat{\gamma} = y_\gamma - X_\gamma (X_\beta' X_\beta)^C X_\beta' y_\beta$ are independent pf the quadratic form $y_\beta' (I_{(n-q)} - X_\beta (X_\beta' X_\beta)^C X_\beta') y_\beta$. To see this, write both forms as functions of y. If $\hat{\gamma}$ is written as B y, then

$$B = \begin{bmatrix} I_q & -X_\gamma (X_\beta' X_\beta)^C X_\beta' \\ 0_{(n-q)xq} & 0_{(n-q)x(n-q)} \end{bmatrix} .$$

If the quadratic form for $\hat{\sigma}^2$ is represented as $\frac{1}{(a-1)(t-1)-q} y' A\, y$, then

$$A = \begin{bmatrix} 0_{qxq} & 0_{qx(n-q)} \\ 0_{(n-q)xq} & I_{(n-q)} - X_\beta (X_\beta' X_\beta)^C X_\beta' \end{bmatrix}$$

It is easily seen that A and B are conformable and that $BA = 0$. Thus, from a well known theorem on the independence of linear and quadratic forms (see Theorem 4.5.1, p. 137, in Graybill[4]) the linear forms representing $\hat{\gamma}$ are independent of the quadratic form $\hat{\sigma}^2$. The result is that, given normal and independent errors, the components of $\hat{\gamma}$ are normally distributed and independent of $\hat{\sigma}^2$ and that t-statistics can be constructed to test for the presence of interaction. Before doing that, the moments of $\hat{\gamma}$ are required.

It is easily seen that $\hat{\gamma}$ is unbiased for γ by substituting (2.6) into (2.5) and using (2.2). Then, $\mathbf{Cov}[\hat{\gamma}] = \mathbf{E}[(\hat{\gamma} - \gamma)(\hat{\gamma} - \gamma)'] = \mathbf{Cov}[By] = \sigma^2 [BB']$, and

$$[BB'] = \begin{bmatrix} I_q + X_\gamma (X_\beta' X_\beta)^C X_\gamma' & 0_{qx(n-q)} \\ 0_{(n-q)xq} & 0_{(n-q)x(n-q)} \end{bmatrix} .$$

Thus,

$$\mathbf{Cov}[\hat{\gamma}] = \sigma^2 (I_q + X_\gamma (X_\beta' X_\beta)^C X_\gamma') .$$

TWO SPECIAL CASES

The algorithm is based on the examination of each cell individually for the presence of interaction. The cells which are believed to contain interaction are simply treated as missing cells in the following ANOVA procedure. The formal estimates are given below.

Suppose that the observations y_{ij} arise from the two-way classification model with one observation per cell, i assumes the values 1, 2, ..., a and j assumes the

values 1, 2, ..., t, and where exactly one of the n = at cells is effected by interaction. Designate that cell as cell (K,L) where $1 \le K \le a$ and $1 \le L \le t$.

The previous results produce normal equations and the parameter estimates. The customary dot notation is used: $y_{i.} = \sum_{j=1}^{t} y_{ij}$ and $\bar{y}_{i.} = \frac{1}{t} y_{i.}$. The superscript O refers to summation within the row or column which contains the cell with interaction: $y^{O}_{K.} = \sum_{j=1,\ j \neq L}^{t} y_{Kj}$; $\bar{y}^{O}_{K.} = \frac{1}{t-1} y^{O}_{K.}$; while $y^{O}_{..}$ refers to the overall sum except for the (K,L)-th cell; $\bar{y}^{O}_{..} = \frac{1}{n-1} y^{O}_{..}$; etc. The estimators are based on the conditional inverse using the usual constraints, $\sum_{i=1}^{a} \alpha_i = \sum_{i=1}^{t} \tau_j = 0$.

$$\hat{\mu} = \bar{y}^{O}_{..} + \frac{1}{t(a-1)} (y^{O}_{K.} - \bar{y}^{O}_{..}) + \frac{1}{a(t-1)} (y^{O}_{.L} - \bar{y}^{O}_{..})$$

$$\hat{\alpha}_k = (\bar{y}_{K.} - \bar{y}^{O}_{..}) + \frac{a-1}{a(t-1)} (\bar{y}^{O}_{.L} - \bar{y}^{O}_{..})$$

$$\hat{\tau}_L = (\bar{y}_{.L} - \bar{y}^{O}_{..}) + \frac{t-1}{t(a-1)} (\bar{y}_{K.} - \bar{y}^{O}_{..})$$

$$\hat{\alpha}_i = (\bar{y}_{i.} - \bar{y}^{O}_{..}) - \frac{1}{t(a-1)} (y^{O}_{K.} - \bar{y}^{O}_{..}) - \frac{1}{a(t-1)} (y^{O}_{.L} - \bar{y}^{O}_{..}) \ ; \ i \neq K$$

$$\hat{\tau}_j = (\bar{y}_{.j} - \bar{y}^{O}_{..}) - \frac{1}{t(a-1)} (y^{O}_{K.} - \bar{y}^{O}_{..}) - \frac{1}{a(t-1)} (y^{O}_{.L} - \bar{y}^{O}_{..}) \ ; \ j \neq L$$

$$\hat{\gamma}_{KL} = (y_{KL} - \bar{y}^{O}_{..}) - \frac{a}{a-1} (\bar{y}_{K.} - \bar{y}^{O}_{..}) - \frac{t}{t-1} (\bar{y}^{O}_{.L} - \bar{y}^{O}_{..})$$

$$\hat{\sigma}^2 = \frac{1}{(a-1)(t-1)-1} \sum_{i,j \neq K,L} (y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\tau}_j)^2$$

The moments of $\hat{\gamma}_{KL}$ are: $E[\hat{\gamma}_{KL}] = \gamma_{KL}$, and $Var[\hat{\gamma}_{KL}] = \frac{at}{(a-1)(t-1)} \sigma^2$.

From the results of part 2, $\hat{\gamma}_{KL}$ and $\hat{\sigma}^2$ are independent so that a test for a non-zero value of γ_{KL} is the student's t with (a-1)(t-1)-1 degrees of freedom. Under the null hypothesis, $\hat{\gamma}_{KL}$ is zero, and the test is two-sided.

$$t = \frac{\hat{\gamma}_{KL}}{\hat{\sigma} \sqrt{\frac{at}{(a-1)(t-1)}}} \ .$$

In the following example, it will be of interest to simultaneously test two cells for significance. Suppose that there are two cells which contain interaction. The estimates differ formally, depending upon whether the two cells are either in the same row or column or in distinct rows and distinct columns.

Suppose the cells are in the same column. Call the cells (K,M) and (L,M). The solution to the normal equations produce the following estimates:

$$\hat{\mu} = \bar{y}^{O}_{..} + \frac{1}{t(a-2)}(y^{O}_{K.} - \bar{y}^{O}_{..}) + \frac{1}{t(a-2)}(y^{O}_{L.} - \bar{y}^{O}_{..}) + \frac{2}{a(t-1)}(y^{O}_{.M} - \bar{y}^{O}_{..})$$

$$\hat{\alpha}_K = (\bar{y}_{K.} - \bar{y}^{O}_{..}) + \frac{a-2}{a(t-1)}(\bar{y}^{O}_{.M} - \bar{y}^{O}_{..})$$

$$\hat{\alpha}_L = (\bar{y}_{L.} - \bar{y}^{O}_{..}) + \frac{a-2}{a(t-1)}(\bar{y}^{O}_{.M} - \bar{y}^{O}_{..})$$

$$\hat{\tau}_M = (\bar{y}_{.M} - \bar{y}^{O}_{..}) + \frac{t-1}{t(a-2)}(\bar{y}_{K.} - \bar{y}^{O}_{..}) + \frac{t-1}{t(a-2)}(\bar{y}_{L.} - \bar{y}^{O}_{..})$$

$$\hat{\alpha}_i = (\bar{y}_{i.} - \bar{y}^{O}_{..}) - \frac{1}{t(a-2)}(y^{O}_{K.} - \bar{y}^{O}_{..}) - \frac{1}{a(t-2)}(y^{O}_{L.} - \bar{y}^{O}_{..}) - \frac{2}{a(t-1)}(y^{O}_{.M} - \bar{y}^{O}_{..})$$

$$\hat{\tau}_j = (\bar{y}_{.j} - \bar{y}^{O}_{..}) - \frac{1}{t(a-2)}(y^{O}_{K.} - \bar{y}^{O}_{..}) - \frac{1}{a(t-2)}(y^{O}_{L.} - \bar{y}^{O}_{..}) - \frac{2}{a(t-1)}(y^{O}_{.M} - \bar{y}^{O}_{..})$$

$$\hat{\gamma}_{KM} = (y_{KM} - \bar{y}^{O}_{..}) - \frac{a-1}{a-2}(\bar{y}_{K.} - \bar{y}^{O}_{..}) - \frac{1}{a-2}(\bar{y}_{L.} - \bar{y}^{O}_{..}) - \frac{t}{t-1}(\bar{y}^{O}_{.M} - \bar{y}^{O}_{..})$$

$$\hat{\gamma}_{LM} = (y_{LM} - \bar{y}^{O}_{..}) - \frac{a-1}{a-2}(\bar{y}_{L.} - \bar{y}^{O}_{..}) - \frac{1}{a-2}(\bar{y}_{K.} - \bar{y}^{O}_{..}) - \frac{t}{t-1}(\bar{y}^{O}_{.M} - \bar{y}^{O}_{..})$$

$$\mathbf{Var}[\hat{\gamma}_{KM}] = \mathbf{Var}[\hat{\gamma}_{LM}] = \frac{t(a-1)}{(t-1)(a-2)}\sigma^2, \text{ and } \mathbf{Cov}[\hat{\gamma}_{KM}, \hat{\gamma}_{LM}] = \frac{t}{(t-1)(a-2)}\sigma^2.$$

If the two cells which contain interaction are in the same row, then the estimates can easily be obtained from the estimates given above by an interchange of indices a and t. The parameter estimates when the two cells occupy distinct rows and columns are much more complicated to write out.

The solutions to the two previous cases and all others are greatly simplified by proceeding as follows. The normal equations are reduced to full rank by first substituting $\hat{\mu}$ into the other equations. Then, the normal equations are arranged so that the coefficients of the cells which contain interaction are listed first beginning on the left, followed by the other coefficients. This produces a "reduced" $X'X$ equation of the following form:

$$R = \begin{bmatrix} A_{qxq} & 0_{qx(a+t-q)} \\ B_{(a+t-q)xq} & I_{(a+t-q)x(a+t-q)} \end{bmatrix}.$$

where A contains only the coefficients of the α_i and the τ_j corresponding to

the cells which contain interaction, 0 contains the coefficients of all other α_i and τ_j, and B is a rank 1 matrix each row of which contains the additive coefficients of $\hat{\mu}$. Then,

$$R^{-1} = \begin{bmatrix} A^{-1} & 0 \\ -BA^{-1} & I \end{bmatrix}.$$

The α_i and τ_j which involve interaction can be solved as a set and then substituted into the equations for the α_i and τ_j. This saves a lot of labor.

AN EXAMPLE

Johnson and Graybill[3] produced an analysis of data in a 3x5 layout with one observation per cell which originally appeared in Black[5]. In their analysis, first Tukey's test is used to determine if, in fact, interaction is present. If it is found to be present, then $k = \binom{a}{2}\binom{t}{2}$ tests for significance among all possible 2x2 contrasts of the cells in the layout are conducted. Apparently then, those cells which are contained in a large number of "significant" contrasts are declared to contain interaction. Those cells are treated as missing observations in a "conventional ANOVA. Since each cell is contained in (a-1)(t-1) of the contrasts, some of the cells which do not contain interaction will appear in some of the "significant" contrasts. Thus, some judgement is involved in deciding which cells are the significant ones. The data in their example is given in table 1, below. In their analysis, cell (1,1) was determined to be significant since all of the 8 = (3-1)(5-1) contrasts which contained that cell produced significant statistics. The contrast for cell (2,1) and cell (3,3) also produced one significant statistic, but that was apparently determined to be due to statistical variation. The mean squared error when cell (1,1) was removed from the analysis was 6,228.34

1. YIELD IN kg/ha OF SPRING WHEAT

Nitrogen in kg/ha	Phosphorous in kg/ha					Total
	0	22	45	90	180	
0	1984	2550	2706	2740	2954	12934
45	1776	2843	3306	3305	3386	14616
90	1797	2761	3240	3227	3332	14357
TOTAL	5557	8154	9252	9272	9672	41907

In our analyis, a t-test was conducted on each cell to determine if interaction is present. No preliminary test for nonadditivity was conducted. The results are given in table 2. The t-statistics were based on 7 degrees of freedom.

2. ANALYSIS OF SPRING WHEAT DATA

Cell	Interaction Estimate	t-statistic	Mean Squared Error
(1,1)	634.999	5.875*	6228.344
(1,2)	73.125	0.279	36542.84
(1,3)	-320.625	-1.372	29117.84
(1,4)	-269.375	-1.109	31421.64
(1,5)	-118.125	-0.455	35887.12
(2,1)	-385.750	-1.760	25612.87
(2,2)	-8.250	-0.031	35945.07
(2,3)	173.625	0.681	34653.43
(2,4)	159.250	0.621	35018.02
(2,5)	61.125	0.233	36665.57
(3,1)	-249.250	-1.014	32216.87
(3,2)	-64.875	-0.248	36629.58
(3,3)	146.999	0.571	35303.84
(3,4)	10.125	0.424	36026.25
(3,5)	57.000	0.217	36702.71

The test indicated that only the (1,1) cell contains interaction which agrees with the original analysis[3]. The statistic 5.875 for cell (1,1) corresponds to a p-value of about 0.0003 with 7 degrees of freedom.

It may be argued that other cells which produce a large (even if not highly significant) statistic should be checked for significance with cell (1,1) removed, since the error estimate would be reduced, resulting in a higher value of the statistic for that cell. In this example, cell (2,1) produced the next highest t-statistic (-1.76) with error estimate 25,612.87 . The error estimate is lower that that produced by the other cells which were in the range 30,000 to 36,000. Then, if cell (1,1) is removed from the analysis, a reduction in the error estimate could produce a significant t-statistic. There is a compensating effect, namely that the removal of the observation in cell (1,1) would reduce the overall average in the first column which would tend to reduce the interaction estimate in cell (2,1).

As an illustration, the special case estimate given in previously is used to produce "revised" estimates for cells (1,1) and (2,1), which are presented in table 3. This is not the recommended procedure because the correlation between cells (1,1) and (2,1) is neglected. A better test is described later.

3. REVISED ESTIMATES OF CELLS (1,1) AND (2,1)

Cell	Interaction Estimate	t-statistic	Mean Squared Error
(1,1)	589.5	4.536*	6755.66
(2,1)	-91.00	-.700	6755.66

This analysis confirms the original conclusions. The estimate for cell (1,1) is highly significant and cell (2,1) is not significant.

As stated previously, a better procedure is to conduct a test which takes full advantage of the correlation between cells (1,1) and (2,1). Based on the results of Section 3, the joint distribution of the estimates $\hat{\gamma}_{11}$ and $\hat{\gamma}_{21}$ are, assuming normal and independent errors, jointly normally distributed with means γ_{11} and γ_{21}, variances $\frac{t(a-1)}{(t-1)(a-2)}\sigma^2$ and covariance $\frac{t}{(t-1)(a-2)}\sigma^2$. Since cell (1,1) has already been judged to contain interaction, the recommended procedure is to test the hypothesis that γ_{21} is zero, given that $\gamma_{11} = 635$ and $\hat{\gamma}_{11} = 589.5$. The appropriate distribution is the distribution of $\hat{\gamma}_{21}$ given γ_{11} and $\hat{\gamma}_{11}$. The conditional distribution of $\hat{\gamma}_{21} \mid (\gamma_{11}, \hat{\gamma}_{11})$ is:

$$N(\gamma_{21} + \frac{1}{a-1}(\hat{\gamma}_{11} - \gamma_{11}) ; \frac{at}{(a-1)(t-1)}\sigma^2) .$$

The appropriate values are: $\hat{\gamma}_{21} = -91$; $\gamma_{11} = 635$; $\hat{\gamma}_{11} = 589.5$; and $\hat{\sigma}^2 = 6{,}755$. These produce the t-value -1.213. The appropriate degrees of freedon are 6. This result agrees with the previous conclusion.

REFERENCES.

[1] Tukey, John W., "One Degree of Freedom for Non-Additivity," Biometrics, 5 (September 1949), 232-42.

[2] Johnson, Dallas E. and Graybill, Franklin A., "An Analysis of a Two-Way Model with Interaction and No Replication" Journal of the American Statistical Association, 67 (December 1972), 862-868.

[3] Johnson, Dallas E. and Graybill, Franklin A., "Estimation of σ^2 in a Two-Way Classification Model with Interaction" Journal of the American Statistical Association, 67 (June 1972), 388-394.

[4] Graybill, Franklin A., Theory and Application of the Linear Model, Belmont Calif. : Wadsworth Publishing Co., 1976

[5] Black, A. L., "Adventitious Roots, Tillers, and Grain Yields of Spring Wheat as Influenced by N-P Fertilization," Agronomy Journal, 62 (January 1970), 32-6.

EXPLORATION OF SATELLITE IMAGES IN THE DYNAMICALLY LINKED ARCVIEW/XGOBI/XPLORE ENVIRONMENT[1]

J. Symanzik*, D. Cook, S. Klinke, and N. Lewin
*George Mason University
Center for Computational Statistics 4A7
Fairfax, VA 22030

ABSTRACT

The dynamic link among the Geographic Information System (GIS) ArcView, the dynamic statistical graphics (DSG) program XGobi, and the statistical computing environment XploRe has been successfully used to explore and analyze all kinds of spatially referenced data — from forest health data over precipitation data to precision agricultural data. In this paper, we will focus on the exploration of satellite images using the linked ArcView/XGobi/XploRe software environment.

INTRODUCTION

Multispectral satellite image data, i. e., remote sensing data gathered through earth observation satellites attracted the attention of statisticians working in the field of dynamic statistical graphics (DSG) during the last decade. Many prototypes of software systems have been developed over time that focus on many interesting software features but yet lack the full functionality that is required to geographically explore satellite images.

The ArcView/XGobi/XploRe software environment consists of the Geographic Information System (GIS) ArcView 3.0™, the dynamic statistical graphics (DSG) program XGobi [25], and the statistical computing environment XploRe [13]. These three programs, linked together into one single environment, provide features such as linked brushing among multiple windows of different types (e. g., map views and statistical plots), grand tour [1], [2], projection pursuit guided tour [6], and standard clustering and classification methods, that are very powerful when jointly used for the exploration and classification of satellite images.

In the following sections we will look at examples from the literature where satellite images have been analyzed using dynamic statistical graphics. We describe the ArcView/XGobi/XploRe software environment, and we explain how it can be used for remote sensing data. We finish with a discussion and an overview on future directions.

SATELLITE IMAGES AND DSG

Various statistical methods have been developed for the classification of remote sensing imagery. Many of these methods are based on principal components and are widely used in the remote sensing community for non–supervised classification purposes. Other statistical approaches that are also well–acknowledged in the remote sensing community have been focused on the optimum band selection for supervised classification. However, most of these solutions ignore the visual capabilities of human beings, highly depend on probably incorrect ground truth, or omit potentially valuable information at an early stage of the analysis.

Exploration of satellite images via DSG is mostly based on human interaction and visual capabilities. This approach helps to detect incorrect ground truth and it does not ignore any of the information that is available. Unfortunately, there exist only experimental software solutions within the statistical community so far that support this approach. Some examples of the exploration of satellite images through DSG follow.

In [4], a Landsat 2 multispectral scanner (MSS) image of the confluence of the Rio Solimões and the Rio Negro near Manaus, Brazil, has been considered. The authors use the Data Viewer [3], [14], a program for the graphical data analysis, to explore the four–dimensional spectral data. The authors are capable to graphically confirm (through linked brushing and the grand tour) that the measurements on a transect across

[1] Approved for public release; distribution is unlimited.
ArcView 3.0 is a trademark of Environmental Systems Research Institute, Inc.

the confluence (mixing) of the two rivers can be represented as a convex combination of the corresponding two water spectra. An earlier video [19] provides more details of this analysis.

In [15], the simple image program MTID and XGobi have been linked together. The authors analyze Landsat Thematic Mapper (TM) images of the agricultural Paraná State region. Based on their graphical analysis, a discrepancy between the images and the ground truth becomes immediately obvious. Probably a shift of the recorded ground truth in relation to the images has occurred.

Scott [23] considers four–channel MSS data from NASA's Landsat IV satellite, taken from an agricultural scene in North Dakota on 5 different days in 1977. Scott displays trivariate scatterplots and contours of the transformed 20–dimensional data and he finds clusters that clearly discriminate between sunflowers and small grains.

Salch and Scott [21] demonstrate a 2–dimensional density grand tour of a Landsat satellite image based on three groups of farm crops. The density grand tour reveals that at least one of the crops contributes to the multimodality of the densities. This is a feature that is also detected in other agricultural settings if one (or more) crop(s) has been planted at different points in time, thus yielding a wide variety of the underlying spectral measurements.

Finally, Carr [5] presents a hexagon bin scatterplot matrix based on the seven spectral bands of a Landsat image at a Nevada test site. This binned scatterplot matrix shows areas of high and low concentrations of numbers of pairs of the seven spectral bands, almost perfect linear relationships between several pairs of bands, a cutoff of intensity values for the thermal band, and bivariate outliers.

Unfortunately, in the last two examples the statistical plots are not directly linked to the original satellite image. Therefore, no immediate geographic conclusion can be drawn from the graphical statistical exploration of the data.

THE ARCVIEW/XGOBI/XPLORE ENVIRONMENT

The ArcView/XGobi/XploRe environment is a software environment where the Geographic Information System (GIS) ArcView 3.0, the dynamic statistical graphics (DSG) program XGobi [25], and the statistical computing environment XploRe [13] have been linked together, thus providing us with three different main concepts in one single environment. We can (1) display spatial locations ans concomitant variables on maps, (2) visualize these concomitant variables using interactively manipulated dynamic statistical graphics, and (3) conduct numerical statistical analyses. While ArcView's strengths are in (1) but it merely supports (2) and (3), XGobi's and XploRe's strengths are in (2) and (3), respectively. A link between these three programs does not only provide the sum of the features of the underlying programs but, in addition, we can gain further insight and understanding of our data through interaction and dynamic linking among the programs. See [27] and [26] for technical details of the software environment and possible applications for spatial data analysis such as real estate data and precision agricultural data.

Remote Procedure Calls (RPCs) are the underlying Interprocess Communication (IPC) mechanism used for the communication among the three programs. More details on this IPC technology can be found in [24] and [9].

In addition to the publications on the entire ArcView/XGobi/XploRe environment, there exist several articles that focus on the preliminary ArcView/XGobi link. A description of the main features of this link (multivariate data, spatial cumulative distribution functions, variogram–cloud plots, spatially lagged scatterplots, and multivariate variogram–cloud plots) and examples that range from forest health data over precipitation data to satellite images can be found in [7] and [8], for example. In [31], a comprehensive summary on the technical details of the ArcView/XGobi link has been given. However, what has been stated there still holds for the entire ArcView/XGobi/XploRe environment. Recent developments on XploRe can be found in [17], [18], and [22].

EXPLORATION OF SATELLITE IMAGES THROUGH ARCVIEW/XGOBI/XPLORE

In this section, we look at three different examples of remote sensing imagery. The first two examples only make use of the ArcView/XGobi link and have been described in more detail in [29] and [30] (Example 1) and [7] (Example 2). Example 3 is based on data presented in [11] and [12] and makes use of the full ArcView/XGobi/XploRe environment.

EXAMPLE 1: LAKE ICARIA WATERSHED

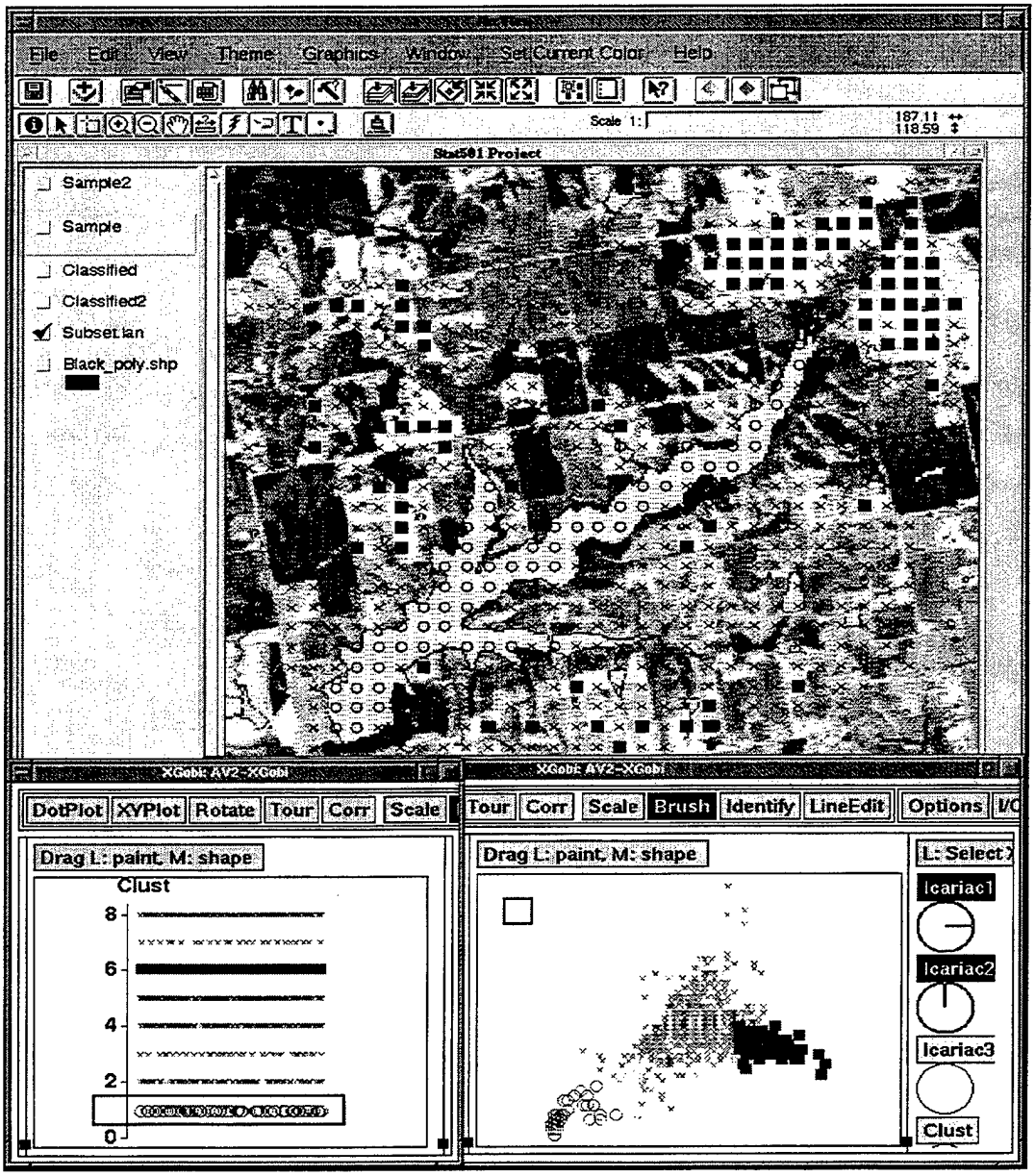


Figure 1: *The ArcView map view (top) is linked with two XGobi views (bottom). The left XGobi shows the 7 clusters obtained from a hierarchical cluster analysis. The right XGobi shows a scatterplot of the variables Icariac1 and Icariac2. The two clusters brushed in the left XGobi fall on lake locations (circles) or areas with active vegetation (filled boxes) in the ArcView map view. They also form two of the corners of the data triangle in the right XGobi.*

This example shows three ranges of wavelengths of electromagnetic radiation (expressed through the variables *Icariac1*, *Icariac2*, and *Icariac3*) measured on April 22, 1990, by the SPOT earth observation satellite. The area under consideration relates to the Lake Icaria watershed in southwest Iowa (near Corning). Instead of using all 300,000 pixels of the scene that are available in ArcView, we took a systematic random sample (e. g., [10], Section 5.6.1) of 800 locations for further analysis in XGobi.

At the time we analyzed this data, XploRe was not yet accessible through a link. It took the user five individual steps to create a new variable *Clust* that contains the result of a hierarchical cluster analysis.

To ease the access to analytical statistical methods we decided to add XploRe to the ArcView/XGobi link. Today, within the ArcView/XGobi/XploRe environment, these steps could be done by pressing a few buttons in ArcView to activate the required functionality within the other two programs.

Figure 1 (a reprint from [29] and [30]) shows the ArcView view (top) and two XGobi views (bottom). We brushed two of the clusters in the dotplot of *Clust* (left XGobi). We see that these clusters are two of the corners of the data triangle in the scatterplot of the variables *Icariac1* and *Icariac2* (right XGobi). In the ArcView view, these points fall on lake locations (circles) or areas with active vegetation (filled boxes). The next step of our interactive analysis might have been the brushing of the third corner of the data triangle in the right XGobi to see if these points also relate to a particular geographic region in the map view. An accompanying video [28] gives additional insight into this interactive and dynamic analysis.

EXAMPLE 2: BORDER REGION BETWEEN VERMONT AND NEW HAMPSHIRE

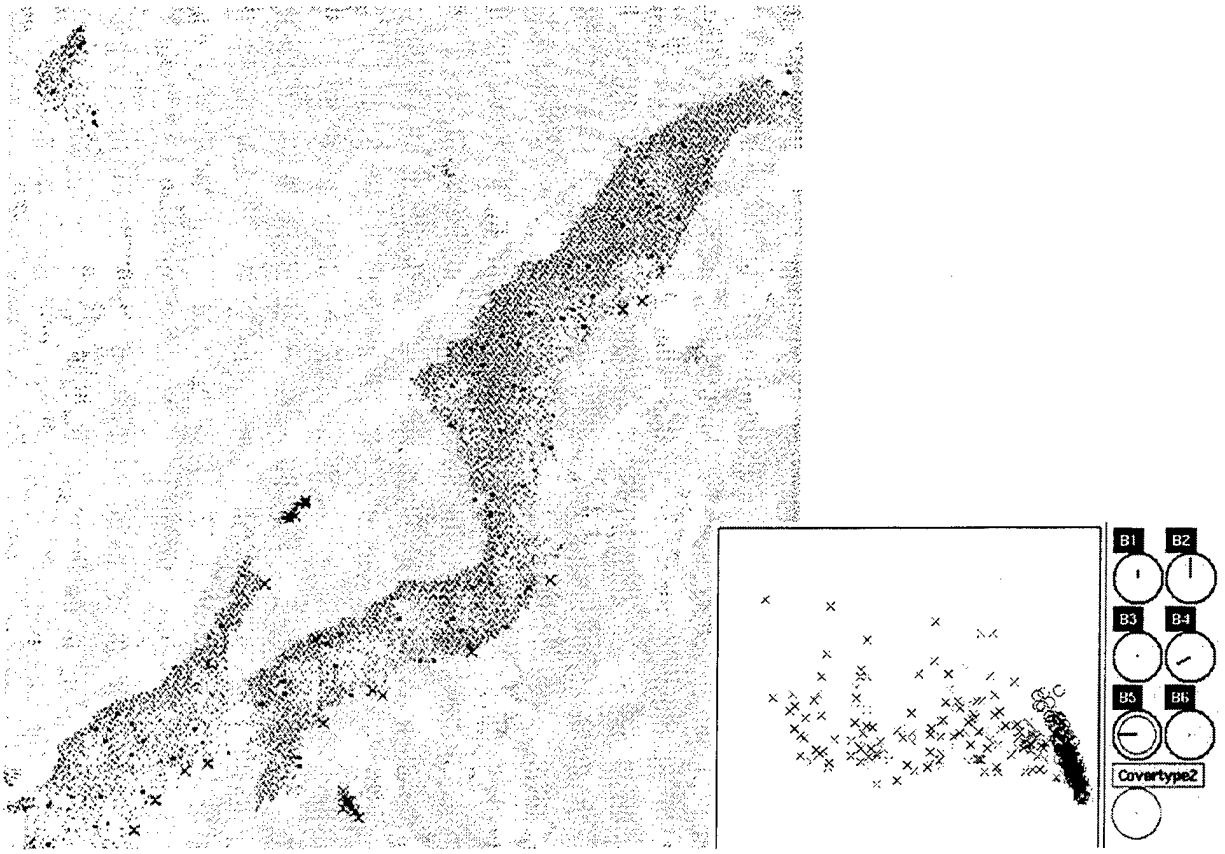


Figure 2: *The ArcView map view (left) shows the spatial region while the XGobi view (right) shows an interesting projection that permits to distinguish between two types of water. The cluster of points (marked with circles) detected in XGobi relates to the middle of the river and the small lake in the upper left corner of the ArcView map view. The remaining points (marked with $\times$'s) that are widely spread in the XGobi view fall onto the smaller ponds and the southern border of the river. The XGobi variable panel indicates that bands 2, 4, and 5 are major contributors to this projection.*

While we only consider main classifications such as water and vegetation in Example 1, this example demonstrates that water is not always equal to water and a quarry can be separated from areas that look similarly in the visible satellite image. This time we have measurements from 6 spectral bands recorded by the Thematic Mapper (TM) instrument of the Landsat earth observation satellite. The area under consideration relates to the border region between the states Vermont and New Hampshire. In Figures 2 and 3 we look at two very small segments of this area. The ArcView view in each of the two figures is based

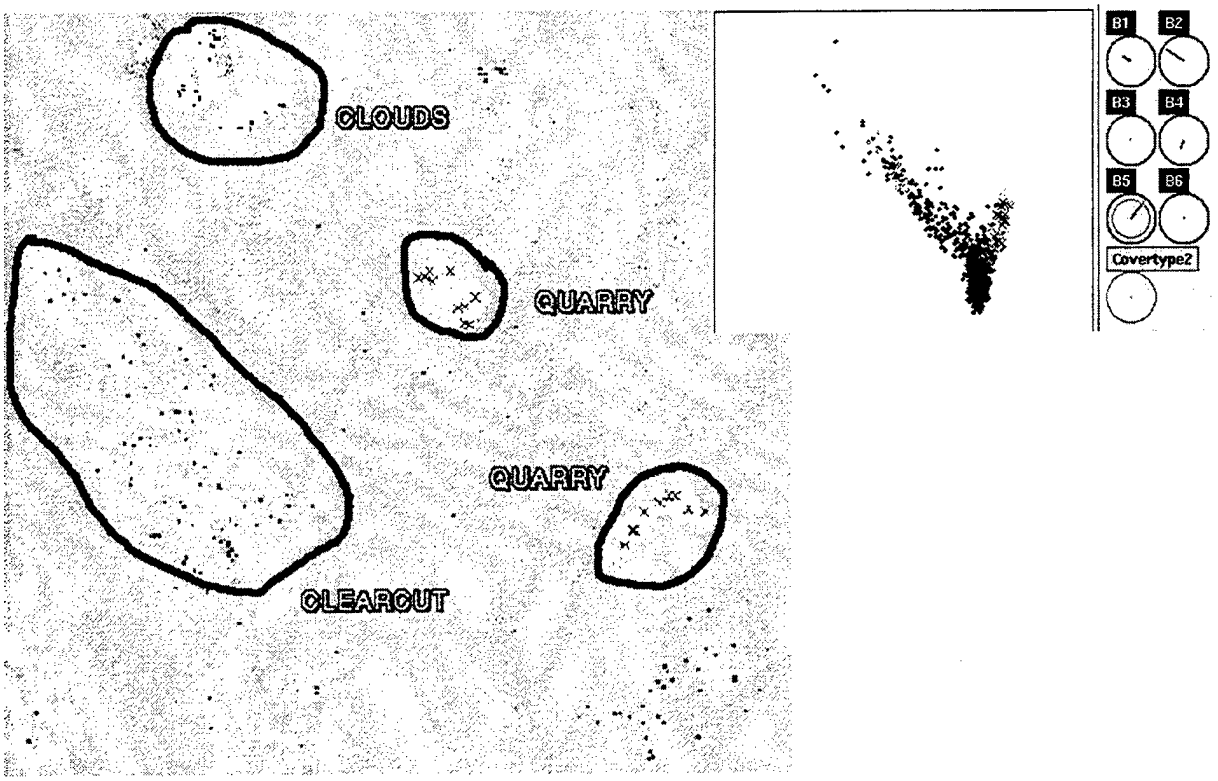


Figure 3: *The ArcView map view (left) shows the spatial region while the XGobi view (right) shows an interesting projection that permits to distinguish among quarry regions (marked with ×'s), a clearcut area, and clouds. The XGobi variable panel indicates that bands 2 and 5 are major contributors to this projection.*

on bands 4, 3, and 2, only. For the XGobi views, we took random samples of points from known origin. The ArcView images have been processed with an image processing program after the analysis to yield a better gray scale representation. On a computer screen you would see red symbols on a mostly green/blue background.

In Figure 2 samples of pixels from the water bodies have been taken. The big water mass that stretches diagonally across the ArcView view (left) relates to the Connecticut river. Three smaller water bodies are also visible. In the XGobi view (right) we see a concentration of circles that are clustered along a line and a large number of ×'s that are scattered in the plot. In the map view we see that the circles fall onto the middle of the river and the small lake in the upper left corner. These points may relate to factors such as a strong current, deep water, and no overgrown vegetation. Otherwise, the ×'s fall onto two smaller ponds and the southern border of the river. These points may relate to shallow water with algae or waterplants in the ponds or overgrown vegetation at the edge of the river. Note that there are no ×'s at the northern border of the river — an indicator that there is no vegetation on this border. Most importantly, the visible ArcView view based on bands 4, 3, and 2 does not allow a distinction among different water bodies. However, the XGobi projection allows such a distinction since, as the XGobi variable panel indicates, band 5 which is not used in ArcView (in addition to bands 2 and 4) is a major contributor to this projection.

In Figure 3 the ArcView view (left) shows a region where 2 clouds, a clearcut area, and two quarry regions have been classified as ground truth. In the XGobi view (right) we can distinguish three regions of points. We brushed the points from the quarry regions with an × in the ArcView view and notice that all these points fall into one of the three regions. However, there are additional points in this region that have not been marked. This might be an indicator for an addional quarry region that has been differently classified in the ground truth. Through brushing in XGobi we would also see where these points are located on the map. Note that once again bands 5 and 2 are major contributors to the XGobi view but band 5 is not a part of the ArcView view.

EXAMPLE 3: AGRICULTURAL SCENE FROM THE IMPERIAL VALLEY

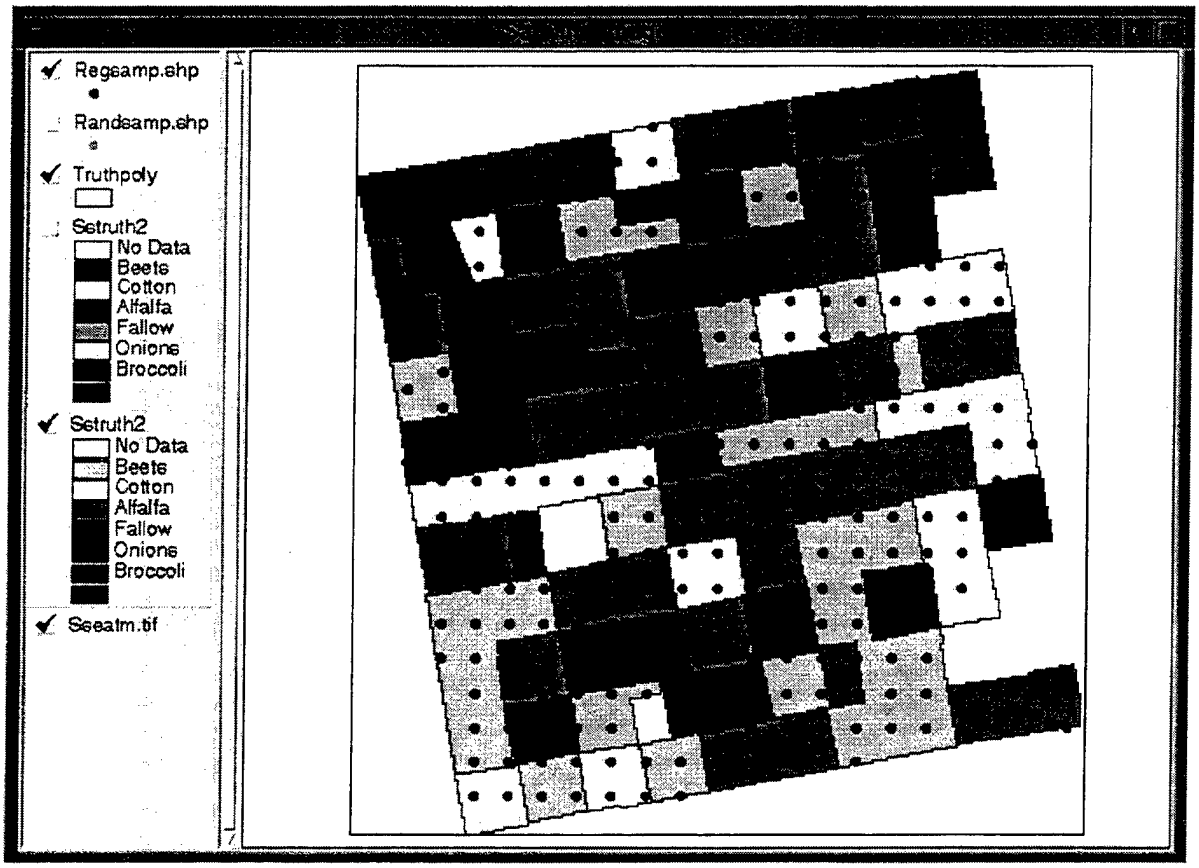


Figure 4: *The ArcView map view shows the reported crop information for an agricultural scene and the sampling locations which have been used for further analysis in XGobi and XploRe. Six crops are identified from the ground. For fields with no crops or mixtures of crops no information has been recorded. There are 7 bands of remotely sensed information available for clustering and comparison with the ground truth.*

In this example we are looking at data from 7 bands recorded by a Landsat–4 TM instrument on December 12, 1982. The scene under consideration represents a very small area of the Imperial Valley in California. In fact, we only consider 124 fields with known crop information provided by the Imperial Valley Irrigation Board. See [11] and [12] for additional details on the data and assessments of different classification techniques. Figure 4 shows the known crop types for the area under investigation. It is a reproduction within ArcView that is based on Figure 2 in [12]. We took a systematic random sample of 314 locations for further analysis in XGobi and XploRe.

As usual, we make use of XGobi's grand tour feature to detect clusters. Figure 5 shows such a cluster in the XGobi view (right). The points that have been brushed (with an open box) fall all into fields of beets in the ArcView map view (left). The ArcView map suggests that some points in the beets fields have been missed through the XGobi classification. This is not the case. As Figures 3, 4, and 5 in [12] indicate, there seem to be streets bordering each field. The points that appear to be located within fields of beets according to the inprecise ground truth used in ArcView most likely fall onto streets, thus yielding quite different wavelength measurements (and therefore projections in the XGobi view) than points from the fields of beets.

Before the next step we deleted points from XGobi and ArcView that have been classified as beets in the previous step, thus we are conducting a hierarchical analysis in this example. Figure 6 shows another cluster (brushed with an open box) in the XGobi view (right) that relates to fields of alfalfa in the ArcView map view (left). However, a large number of alfalfa points in the map is not highlighted. [12], again, provides an explanation: There are "new" and "old" alfalfa fields. Unfortunately, this additional information was not

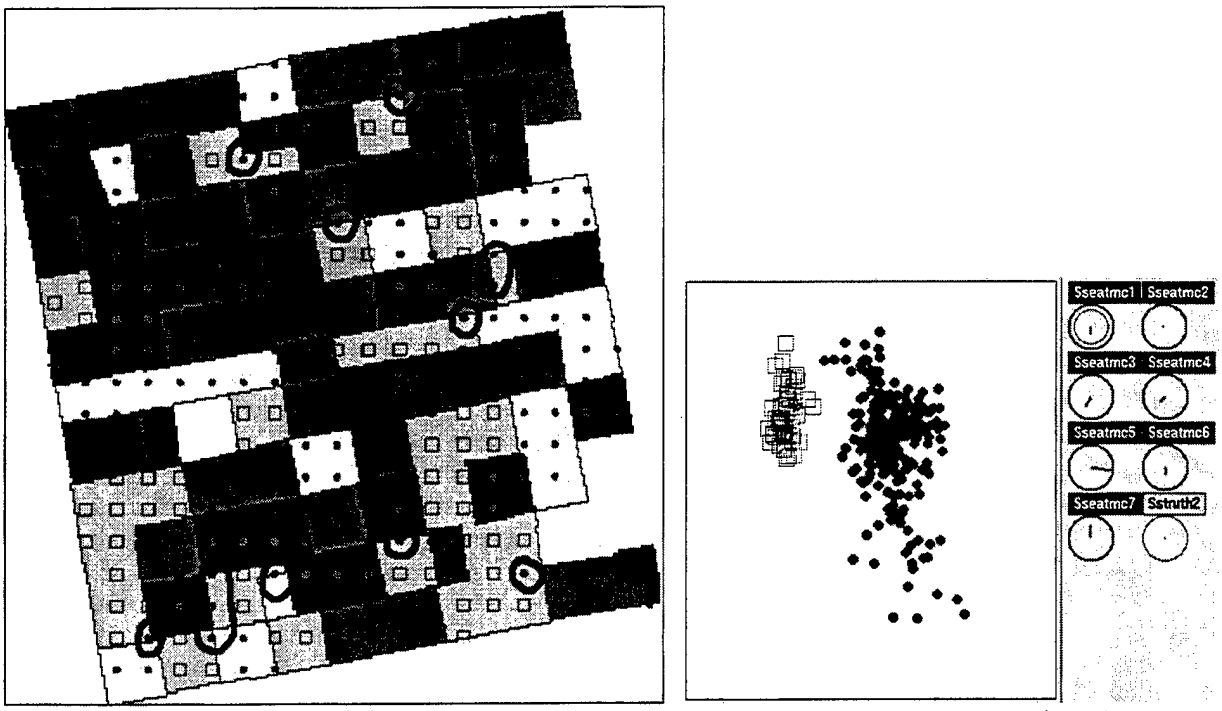


Figure 5: *The ArcView map view (left) shows the spatial region while the XGobi view (right) shows an interesting projection that permits to distinguish between beets (open boxes) and other crops (filled circles). Points that appear to fall on beets fields but are not contained in the XGobi cluster have been circled in the ArcView map view. From additional sources we know that these points most likely fall onto streets dividing or surrounding the fields.*

available for our analysis. However, it seems to be very likely that our given projection in XGobi allows a distinction between "new" and "old" alfalfa fields.

One should note that in Figure 6 we also brushed the single sampling point that falls into the broccoli field. In the XGobi view the corresponding filled box appears at one edge of the alfalfa cluster. It would be interesting to have a larger sample of points from the broccoli field and see if a distinction between alfalfa and broccoli is possible.

In our interactive environment we would try to find additional clusters in XGobi that allow to distinguish among the other type of alfalfa, cotton fields, and fallow fields. Since we have only two sample points for onions, it is very likely that these will not be identified as an additional crop. Taking more sample points from this crop's field would probably help in clustering it.

If we look at the XGobi variable panels in Figures 5 and 6, we see that each of the 7 bands has a non–neglectable influence on at least one of the visible projections. Any classification approach that tries to eliminate any of the bands will most likely produce a less precise classification.

In addition to the visual approach in XGobi presented so far we make also use of XploRe's clustering capabilities to analyze this data set. XploRe supports two basic clustering methods, i. e., hierarchical clustering and partitioning clustering. Partitioning methods (`kmeans` in XploRe) require an initial partition. Other k–means clustering methods, like adaptive k–means clustering, are also available in XploRe.

In this example we make use of a hierarchical method. For this approach interpoint distances of cluster centers will be used to build additional clusters. We start with n clusters that contain one observation each and successively merge clusters together. XploRe allows different hierarchical methods like single linkage, complete linkage, average linkage, Centroid linkage, Ward method, and Lance– and Willams method. For a short overview of these methods see [16], for a more detailed overview see [20], and for an overview of methods implemented in XploRe see [13].

To use the hierarchical method we have to select a distance measure and one of the previously listed methods. We choose the Euclidean distance and the Ward method. The Ward method minimizes the

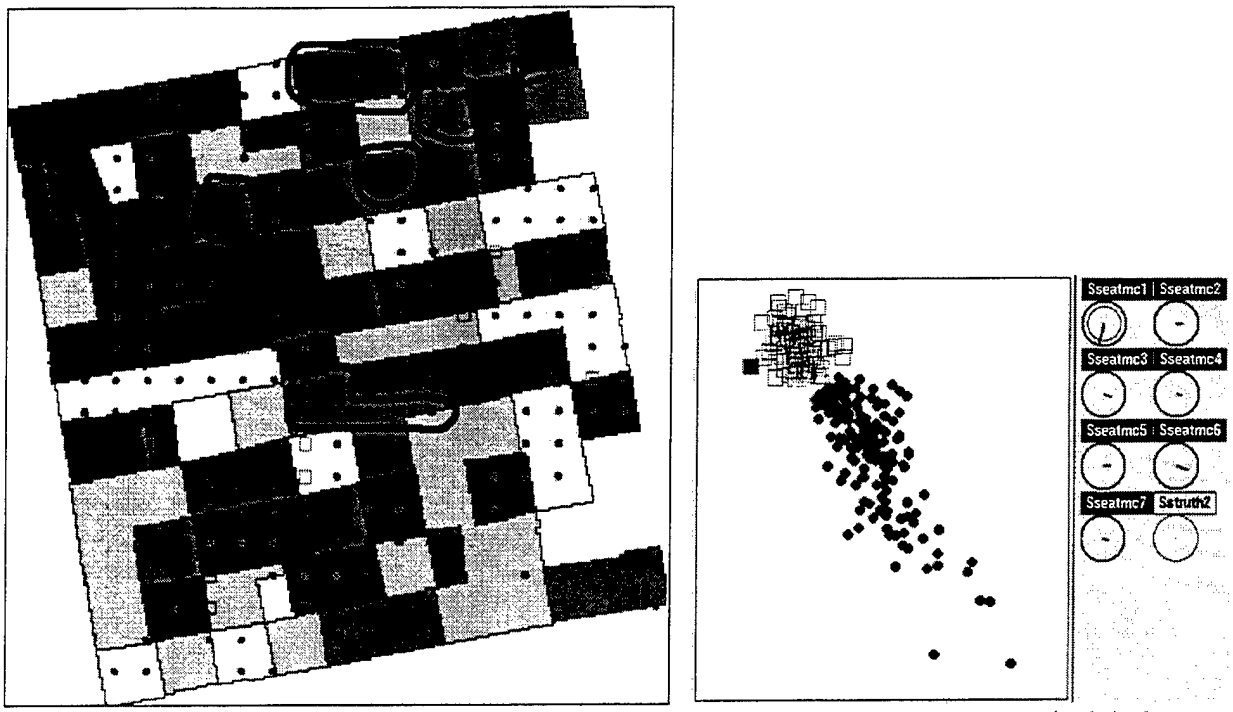


Figure 6: *The ArcView map view (left) shows the spatial region while the XGobi view (right) shows an interesting projection that permits to distinguish between alfalfa (open boxes) and other crops (filled circles). Alfalfa fields that are not contained in the XGobi cluster have been circled in the ArcView map view. There are "new" and "old" alfalfa fields. The cluster in XGobi most likely relates to only one of these types. Broccoli (just one sample point represented by the filled box) has values similar to alfalfa.*

"within" cluster variances. The dendrograms in Figure 7 show the logarithm of the sum of the "within" cluster variances on the y–axis. We see at which levels observations and clusters are merged together (bottom: all 313 observations[2], top: one cluster). Since the variances of the clusters vary, we applied clustering on the original data and on the standardized data. Both dendrograms show a large increase of the sum of the "within" cluster variances if we merge the last two clusters. The dendrogram of the standardized variables shows also an increase of the sum of the "within" cluster variances if we have four clusters.

Since we know the clustering based on ground truth, we can compare the results of the clustering algorithm (based on the original data) and the true groups as in Table 1. The Ward methods behave a little bit worse on the standardized data (not shown). However, Table 1 shows that with two clusters (left part) just group 1 (beets) is somewhat separated from the other crops. An increase to four clusters (right part) shows that group 3 (alfalfa) is split into three of the four clusters. This is not surprising since we already know that there are different types of alfalfa. Group 1 (beets) is separated very well from the other crops. The first cluster does not only contain parts of group 3 (alfalfa) but also groups 2 (cotton) and 4 (fallow). Obviously, much more than four clusters are required to distinguish among these crops. This matches the results of [11] where a clustering method based on unsupervised signatures resulted in 19 clusters. Since groups 5 (onions) and 6 (broccoli) contain only 2 and 1 observations, respectively, it is basically impossible to detect these as separate clusters.

Figure 8 shows some good bivariate projections based on the partition found in the data. The two cluster projection (left) shows a strong linearity within the data whereas the four cluster projection (right) shows how well the clusters are separated in the data.

[2]From the dataset one observation was deleted since the measurement of all channels was zero.

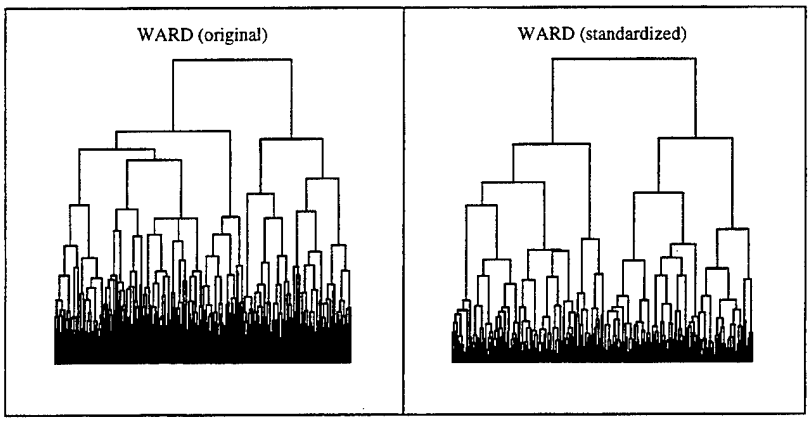


Figure 7: *The dendrograms for the original data (left) and standardized data (right) indicate how clusters are merged together. At the bottom we have all 313 observations and on top we have one cluster. On the y–axis we see the logarithm of the sum of the "within" cluster variances.*

Table 1: *Comparison of clustering (based on the original data) and the ground truth based on two clusters (left part) and four clusters (right part). The numbers 1 to 6 on top of the table stand for the 6 crops, i. e., beets, cotton, alfalfa, fallow, onions, and broccoli.*

```
[ 1,]
[ 2,]
[ 3,] Crosstable for variables 1, 2
[ 4,]
[ 5,]         |      1.0000  2.0000  3.0000  4.0000  5.0000  6.0000 |
[ 6,] --------|-------------------------------------------------------|--------
[ 7,]  1.0000 |      9       47      70      72      2       0      |    200
[ 8,]  2.0000 |      58      0       52      2       0       1      |    113
[ 9,] --------|-------------------------------------------------------|--------
[10,]         |      67      47      122     74      2       1      |    313
[11,]
[12,] Chi^2 test of independence
[13,]
[14,]   chi^2 statistic:                    141.45
[15,]   degrees of freedom:                   5
[16,]   significance level for rejection:   0.0000
[17,]
[18,]   contingency coefficient:             0.56
[19,]   corrected contingency coefficient:   0.79
```

```
[ 1,]
[ 2,]
[ 3,] Crosstable for variables 1, 2
[ 4,]
[ 5,]         |      1.0000  2.0000  3.0000  4.0000  5.0000  6.0000 |
[ 6,] --------|-------------------------------------------------------|--------
[ 7,]  1.0000 |      4       45      35      71      2       0      |    157
[ 8,]  2.0000 |      5       2       35      1       0       0      |    43
[ 9,]  3.0000 |      1       0       51      0       0       0      |    52
[10,]  4.0000 |      57      0       1       2       0       1      |    61
[11,] --------|-------------------------------------------------------|--------
[12,]         |      67      47      122     74      2       1      |    313
[13,]
[14,] Chi^2 test of independence
[15,]
[16,]   chi^2 statistic:                    392.40
[17,]   degrees of freedom:                  15
[18,]   significance level for rejection:   0.0000
[19,]
[20,]   contingency coefficient:             0.75
[21,]   corrected contingency coefficient:   0.86
```

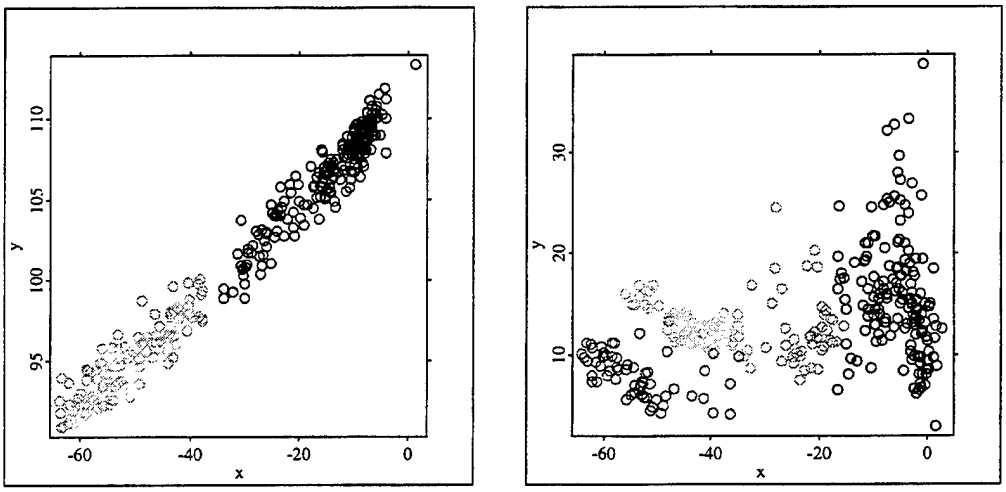


Figure 8: *Bivariate projections based on the partition found in the data. The two cluster projection (left) shows a strong linearity within the data whereas the four cluster projection (right) shows how well the clusters are separated in the data.*

SUMMARY AND FUTURE DIRECTIONS

In this paper we described the use of the ArcView/XGobi/XploRe software environment for the exploration and analysis of satellite images. Within this environment we are capable to detect interesting features of the remote sensing imagery due to its appearance in a series of dynamic plots in XGobi. We can also conduct an analytical statistical analysis of the data in XploRe.

Further work can be directed towards two directions. First, one should consider how existing classification techniques of remote sensing imagery can be combined with the visual approach highlighted in this paper. What is the gain if we combine these methods, how many additional percent of pixels do we classify correctly?

The work presented in this paper should not be misunderstood. We do not claim at all that the classification through dynamic statistical graphics presented here is better than any existing classification technique. But we are convinced that existing classification techniques combined with the graphical approach help to produce better classification results in an easy way. This should be further investigated.

Direction two deals with possibilities for improvements of the software environment. For example, it is absolutely necessary to solve one restriction of this environment, i. e., the fact that only one of the graphical links is fully functional at a time. For example, if we brush points in the XGobi view, they are also brushed either in ArcView or in XploRe — but not in both other programs at the same time what one would expect from such a linked environment. A solution has been designed (see the part on "hierarchical" linked brushing in [31] and [26]) but it has not been implemented yet.

There exist other possible extensions that would be particularly useful for the exploration of satellite images. As we have seen in all of our examples, we typically only use a very small subset of the pixels of a satellite image. So far, we used individually written AVENUE scripts in ArcView to do this sampling, e. g., systematic random sampling or stratified sampling based on ground truth such as water, road, or urban surfaces. An easily accessible collection of sampling mechanisms in this environment would be ideal.

Also the concepts of "geographic brushing" through ArcView (e. g., brush all points that are at most 1 km away from a known water source) and "statistical brushing" through XploRe (e. g., brush all points where the greenness index — a function based on measurements of some of the satellite bands — falls into a given range to show actively growing vegatation) would be additional useful extensions of this software environment for the exploration of satellite images and other geographically referenced data.

ACKNOWLEDGMENTS

Symanzik's work was partially supported by a National Science Foundation Group Infrastructure Grant DMS–9631351. Additionally, it was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 373 "Quantifikation und Simulation ökonomischer Prozesse", Humboldt–Universität zu Berlin. The article has not been subjected to the review of any of the previously mentioned agencies and thus does not necessarily reflect the view of any of these agencies and no official endorsement should be inferred. We would like to thank Barry Haack and Matthew Bechdol for providing us with the satellite image used for Example 3.

REFERENCES

[1] D. Asimov. The Grand Tour: A Tool for Viewing Multidimensional Data. *SIAM Journal on Scientific and Statistical Computing*, 6(1):128–143, 1985.

[2] A. Buja and D. Asimov. Grand Tour Methods: An Outline. In D. M. Allen, editor, *Proceedings of the 17th Symposium on the Interface between Computer Science and Statistics, Lexington, KY*, pages 63–67. Elsevier, 1986.

[3] A. Buja, C. Hurley, and J. A. McDonald. A Data Viewer for Multivariate Data. In T. J. Boardman and I. M. Stefanski, editors, *Proceedings of the 18th Symposium on the Interface between Computer Science and Statistics, Fort Collins, CO*, pages 171–174. American Statistical Association, Washington, D.C., 1986.

[4] A. Buja, J. A. McDonald, J. Michalak, and W. Stuetzle. Interactive Data Visualization Using Focusing and Linking. In G. M. Nielson and L. Rosenblum, editors, *Proceedings of Visualization '91, Los Alamitos, CA*, pages 156–163. IEEE Computer Society Press, 1991.

[5] D. B. Carr. Looking at Large Data Sets Using Binned Data Plots. In A. Buja and P. T. Tukey, editors, *Computing and Graphics in Statistics*, pages 7–39. Springer, New York, NY, 1991.

[6] D. Cook, A. Buja, J. Cabrera, and C. Hurley. Grand Tour and Projection Pursuit. *Journal of Computational and Graphical Statistics*, 4(3):155–172, 1995.

[7] D. Cook, J. J. Majure, J. Symanzik, and N. Cressie. Dynamic Graphics in a GIS: Exploring and Analyzing Multivariate Spatial Data Using Linked Software. *Computational Statistics*, 11(4):467–480, 1996.

[8] D. Cook, J. Symanzik, J. J. Majure, and N. Cressie. Dynamic Graphics in a GIS: More Examples Using Linked Software. *Computers and Geosciences: Special Issue on Exploratory Cartographic Visualization*, 23(4):371–385, 1997. Paper, CD, and `http://www.elsevier.nl/locate/cgvis`.

[9] J. R. Corbin. *The Art of Distributed Applications: Programming Techniques for Remote Procedure Calls.* Springer, New York, Berlin, Heidelberg, 1991.

[10] N. A. C. Cressie. *Statistics for Spatial Data (Revised Edition).* Wiley, New York, NY, 1993.

[11] B. Haack and S. Jampoler. Agricultural Classification Comparisons Using Landsat Thematic Mapper Data. *ITC Journal*, 1994(2):113–118, 1994.

[12] B. Haack and S. Jampoler. Colour Composite Comparisons for Agricultural Assessments. *International Journal on Remote Sensing*, 16(9):1589–1598, 1995.

[13] W. Härdle, S. Klinke, and B. A. Turlach. *XploRe: An Interactive Statistical Computing Environment.* Springer, New York, Berlin, Heidelberg, 1995.

[14] C. Hurley. A Demonstration of the Data Viewer. In E. J. Wegman, D. T. Gantz, and J. J. Miller, editors, *Proceedings of the 20th Symposium on the Interface between Computing Science and Statistics, Fairfax, VA*, pages 108–114. American Statistical Association, Alexandria, VA, 1988.

[15] R. Klein and R. I. Moreira. Exploratory Analysis of Agricultural Images via Dynamic Graphics. Technical Report 9/94, Laboratório Nacional de Computação Científica, Rio de Janeiro, Brazil, 1994.

[16] S. Klinke. *Data Structures for Computational Statistics.* Physica–Verlag, Heidelberg, 1997.

[17] S. Klinke and T. Kötter. XploRe 4 — A Statistical Computing Environment. In F. Faulbaum and W. Bandilla, editors, *SoftStat '95 Advances in Statistical Software 5*, pages 113–122, Stuttgart, 1996. Lucius & Lucius.

[18] T. Kötter. Development of XploRe 4: The Programming Language. In A. Prat and E. Ripoll, editors, *Compstat – Proceedings in Computational Statistics, Short Communications and Posters*, pages 201–202, 1996.

[19] J. A. McDonald and S. Willis. Use of the Grand Tour in Remote Sensing. ASA Statistical Graphics Video Lending Library (contact: dfs@research.att.com), 1987.

[20] H. J. Mucha. *Clusteranalysis auf Mikrocomputern.* Akademie Verlag, Berlin, 1992.

[21] J. D. Salch and D. W. Scott. Data Exploration with the Density Grand Tour. *Statistical Computing and Statistical Graphics Newsletter*, 8(1):7–11, 1997.

[22] S. Schmelzer, T. Kötter, S. Klinke, and W. Härdle. A New Generation of a Statistical Computing Environment on the Net. In A. Prat, editor, *Compstat – Proceedings in Computational Statistics*, pages 135–148, Heidelberg, 1996. Physica–Verlag.

[23] D. W. Scott. Data Analysis in Three and Four Dimensions with Nonparametric Density Estimation. In E. J. Wegman and D. J. DePriest, editors, *Statistical Image Processing and Graphics*, pages 291–305. Marcel Dekker, New York, NY, 1986.

[24] W. R. Stevens. *UNIX Network Programming.* Prentice–Hall, Englewood Cliffs, NJ, 1990.

[25] D. F. Swayne, D. Cook, and A. Buja. XGobi: Interactive Dynamic Graphics in the X Window System. *Journal of Computational and Graphical Statistics*, 1998. Forthcoming.

[26] J. Symanzik, S. Klinke, S. Schmelzer, D. Cook, and N. Lewin. The ArcView/XGobi/XploRe Environment: Technical Details and Applications for Spatial Data Analysis. *ASA Proceedings of the Section on Statistical Graphics*, 1997. Forthcoming.

[27] J. Symanzik, T. Kötter, S. Schmelzer, S. Klinke, D. Cook, and D. Swayne. Spatial Data Analysis in the Dynamically Linked ArcView/XGobi/XploRe Environment. *Computing Science and Statistics*, 29, 1997. Forthcoming.

[28] J. Symanzik, J. J. Majure, and D. Cook. Dynamic Graphics in a GIS: Analyzing and Exploring Multivariate Spatial Data. ASA Statistical Graphics Video Lending Library (contact: dfs@research.att.com), 1995.

[29] J. Symanzik, J. J. Majure, and D. Cook. Dynamic Graphics in a GIS: A Bidirectional Link between ArcView 2.0 and XGobi. *Computing Science and Statistics*, 27:299–303, 1996.

[30] J. Symanzik, J. J. Majure, and D. Cook. Dynamic Graphics in a GIS: A Bidirectional Link between ArcView 2.1 and XGobi — An Update. *Computing Science and Statistics*, 29(2):35–40, 1997.

[31] J. Symanzik, J. J. Majure, D. Cook, and I. Megretskaia. Linking ArcView 3.0 and XGobi: Insight Behind the Front End. Technical Report 97–10, Department of Statistics, Iowa State University, Ames, Iowa, 1997.

VISUAL EXPLORATION OF SPATIAL DATA WITH MANET

Adalbert F.X. Wilhelm*
Center for Computational Statistics
George Mason University
Fairfax, Virginia 22030

ABSTRACT

MANET is research software developed at the Department for Computeroriented Statistics and Data Analysis at the University of Augsburg. MANET offers visual exploration of data sets. It is based on the paradigm of linking low-dimensional views in a highly interactive environment. Two main features of MANET are the consistent treatment of missing data in visualization and the link between graphic representations of geographic space and the graphic representations of attribute space. This paper shows various examples how this link can be used to explore data with underlying geographic information.

INTRODUCTION

Spatial data are widely used to demonstrate new developments in interactive statistical graphics. This is not a mere coincidence but the consequence of the extent to which spatial data and interactive graphics can supplement each other. Spatial data are often characterized as being "data rich but theory poor"[1] and they often do not meet the standard assumptions necessary to perform a confirmatory analysis. Interactive graphics are based on the philosophy of exploratory data analysis (EDA) established by Tukey and his scholars. In EDA the data should speak for themselves or in Tukey's words "EDA is looking at data to see what it seems to say"[2]. The main goal in EDA is to detect patterns, to produce hypotheses and to identify potentially misleading or influential observations, such as outliers or leverage points. All these premises fit well to the analysis of spatial data. In addition, geographers are used to work with graphical representations of their data, like maps and cartograms.

It is therefore not surprising that a series of work has been done that tries to connect interactive statistical graphics and spatial data analysis. Three different branches of software developments accompany these efforts: integrating spatial ideas in interactive statistical software, integrating interactive graphics in spatial analysis software and linking Geographic Information Systems to dynamic graphics software. Examples for the latter are the XGobi/ArcView Link[3] and the SpaceStat/ArcView Link[4]. The second approach is pursued in the development of CDV[5], the first one in REGARD[6] (formerly SPIDER)[7] and in MANET[8].

Typically, integrated systems are considerably faster in execution than coupling approaches and they provide a more seamless environment which is so crucial for a highly interactive exploration.

MANET has three components with a mainly spatial operational field (the polygon map, the trace function, and weighted plots) and it is strongly influenced by the experiences made with REGARD.

Interactive statistical graphics resulted from the merge of static data representation used for exploratory data analysis and inexpensive graphics-capable desktop computers. But interactive graphics is much more than just creating a graphic on the computer screen. The possibility of user interaction with instantaneous response opens a completely new way of looking at data. Whereas the goal in cartography and static graphics is to show in one plot as much information as possible, interactive graphics keep the displays simple by hiding information that is not essential for interpretation until it is specifically requested.

As yet, there does not exist a precise definition of interactive statistical graphics. Often researchers use the notions dynamic graphics, direct manipulation graphics and interactive graphics interchangeably. Eick and Wills[9] define "*Interactive Graphical Methods* as the class of techniques for exploring data that allow the user to manipulate a graphical representation of the data". Moreover they define "an *Interactive*

*On leave from Universität Augsburg, Germany.

Graphic View as a pictorial representation of some form of data or information which the analyst can manipulate in real time"[9]. Current statistical software packages offer at least scatterplot brushing and 3-D-rotation as the main interactive procedures, but a much bigger variety of features has been proposed in the literature and made available as prototypes. Interactive graphics are indispensable for acquisition of qualitative insights into the data sets, for studying model residuals as well as for revealing quantitative results.

The principal concepts to make a single graphic interactive are:

Changing Projection Views Paper and screen are restricted to two dimensions and the human eye-system is trained for the three-dimensional world only. Dimension reduction techniques are used to guarantee quick and easy perception. To regain some of the multi-dimensionality of the data set a rapid dynamic and smooth change of the projection views must be provided.

Rescaling Perception of graphical displays strongly depends on the scale. Since there are no unique choices, statistical software should provide the user with tools to flexibly change plot scales.

Interrogation Graphics should not be overloaded. On demand additional information must be available directly from graphics.

Selection Selecting subgroups and focusing on specific data points helps to reveal structure in the data set. A wide variety of tools to select groups of points from graphical representations is needed to perform a sophisticated analysis.

Linking Full interactivity is only achieved when selection is not restricted to a single display but propagated to other plots. This means that all displays are connected and that each view of the data shows each case consistently. Linking is the key concept of interactive statistical graphics, it builds up a relation between measurements of various variables, between different graphical representations as well as between raw data and some models. These links can also perform different functions - the standard one is highlighting, color encoding or hiding are others.

For all these interactions an instantaneous response is crucial. As Eick and Wills point out this means that a "response time of 50ms or less is required"[9].

Two different conceptual approaches exist for interactive statistical graphics: single window linking and multiple window linking. The first concept requires that all linked plots are arranged in one single window. Therefore, linking is often unnecessarily restricted to array arrangements of the same plot type, e.g. scatterplot matrices and trellis displays[10]. In the multiple window approach each display lives independently, but they all inherent selection information from the case labels.

For spatial data the paradigm of multiple windows linking is very appealing. It seems quite natural to combine the geographic location in form of a map to the graphical representations of the variable measurements. But in fact, maps have only been added quite recently as an additional view of the data, see Monmonier[1]1, Haslett et al 7 and MacDougall[1]2.

Linking low-dimensional views is particularly helpful for the standard questions of exploratory data analysis, like searching for outliers and clusters, or investigating distributional properties like symmetry or modes. In contrast to high-dimensional rotating point clouds or parallel coordinate plots, linking low-dimensional views allows to use displays like histograms and boxplots that convey the marginal distributional properties much better. Interpretation of linked views is typically much easier than for complicated high-dimensional plots since the user controls the process of investigation. But still a lot of the multivariate structure can be found.

INTERACTIVE MAPS

The key tool to enhance interactive graphics software to the analysis of spatial data is to add a map tool. Data with point locations can simply be handled by drawing a scatterplot for longitude and latitude, but many spatial data sets are based on regions. MANET is able to deal with such data by drawing a polygon plot. How can we make a polygon map interactive? Changing projection views could be interpreted as using different projection techniques to map the threedimensional locations into planar coordinates. A less sophisticated approach is to allow different representations of the same map.

This is done in MANET where the user can switch between a filled polygon map, a hollow polygon map and a choropleth map, see Figure 1. Switching from one representation to another is easy and can be performed by simple mouse clicks. The filled representation is more effective in showing the global spatial

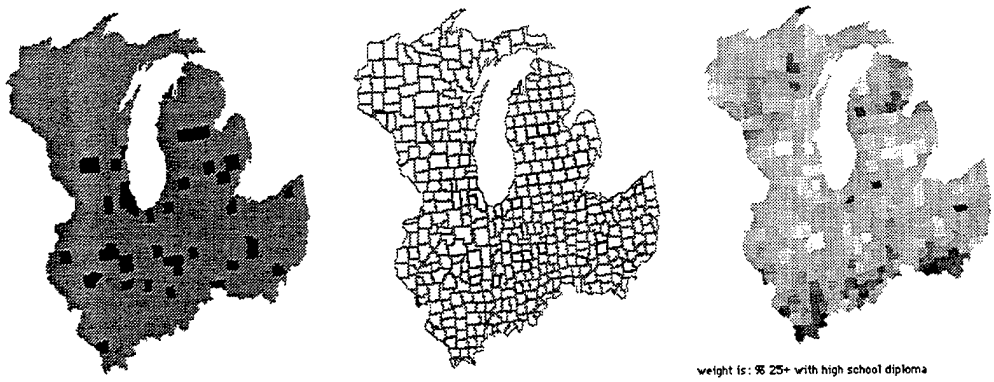


Figure 1: The default polygon map is a filled one (left). Changing plot parameters draws only the boundaries (center). Dragging and dropping a variable over the plot changes it to a choropleth map (right).

distribution but it is highly affected by the size of the regions. Perception of spatial distributions shown in choropleth maps depends heavily on the scale used. In MANET two sliders are available to transform the data towards a more appropriate scaling. A preview of the color distribution makes it easy to find a transformation that results in a more informative choropleth map, see Figure 2. The distraction is here mainly caused by one outlier with a multi-million population: Chicago – that shows up as a white spot. The transformed map still shows Chicago as an outlier, but it shows the variability in the bulk of the data much better. In addition, a second outlier appears at the lower end of the population scale: Keweenaw, the county that is farthest north.

Transformations can also be used to reverse the color table. The default assigns white for the highest values and black for the lowest, but depending on the context interpretation might be easier or the display might be more consistent with a reversed color table.

The choropleth map is not restricted to continuous variables. Discrete variables are handled in the same way resulting in what is often called a chorochromatic or k-color map.

MANET treats the map as any other graphical data display. Therefore, all facilities for user interaction provided for statistical graphs, such as interrogation, are also available for the map.

LINKING MAPS AND STATISTICAL GRAPHICS

Scatterplot brushing is today a fairly well-known exploratory visual technique. But the idea of brushing is not restricted to scatterplot matrices only. There are quite a few software packages available that offer brushing for all kind of statistical graphics, like bar charts, histograms or box plots. From this point of view it is straightforward to link the geographic information displayed in a map to the attribute information displayed in a graphic. In MANET the polygon map is fully linked to all attribute displays. For any subsets of areas selected on the polygon map the distributions of the corresponding observations are highlighted in all non-spatial views. Similarly, for any subset of data selected in a statistical graph, such as a category in a bar chart, the corresponding areas are highlighted on the map.

The basic application of that feature is to focus on interesting subsets and to compare their characteristics with either the entire data set or the complementary subset. Methodologically, there is no difference whether the selection is made on the map or on a statistical graph. In Figure 3 counties in the Midwest region in which a high proportion of elderly people is known to be poor are selected (see boxplot on the right). Immediately, the same counties are highlighted in all other plots. The boxplots for the age categories 0-17 and 18-59 show that poverty is – as expected – a family related problem that affects all generations. More surprisingly, however, is the fact that poverty has a specific spatial distribution. All the selected counties lie in the south of our region of investigation.

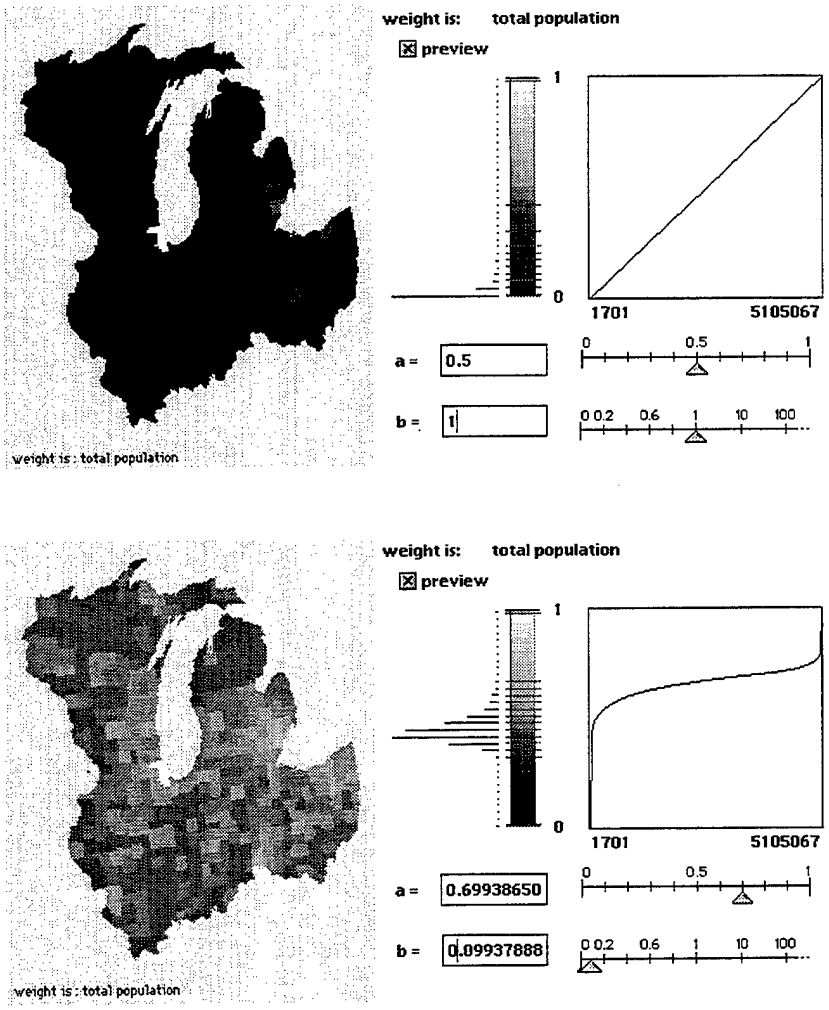


Figure 2: On the top is the default choropleth map and the corresponding frequency distribution. Changing the sliders in the preview leads to an improved map shown on the bottom left.

Subset selections can also be based on both geographic and attributive information at the same time. In MANET a series of selections can be composed with logical operations and all steps of such sequences are stored and can be changed individually. Thus, quite elaborate data queries can be performed graphically and various subgroups can be checked effectively[13].

Tracing is a special kind of generalized brushing proposed by Craig, Haslett, Unwin and Wills[14]. It computes statistics like the mean, the span, or the standard deviation, of specified variables for all points covered by the brush rectangle. As the brush moves over the window the computations are updated and the results are displayed in a new view in a time series plot format (see Figure 4). Tracing is particularly useful for classifying variables and detecting spatial dependence. Linking in trace plots does not reflect a one-to-one correspondence but either a one-to-m or a m-to-m correspondence. Selecting a county in the map will highlight all calculations in the trace plot that use the county's measurements. Selecting a point in the trace plot will highlight all counties in the map that contribute to this value. In addition, all other points in the trace plot that also make use of the highlighted counties' measurements are selected.

Spatially referenced area data is typically based on politically defined regions. Bar charts or histograms that reflect the number of regions falling in a particular class are often misleading and bear low information. Instead, some demographic figures, like total population or population density, might be more appropriate to use as weight for each bin area than the bare number of regions. In MANET weighting is possible for bar charts, histograms and mosaic plots and it is performed by simply multiplying each

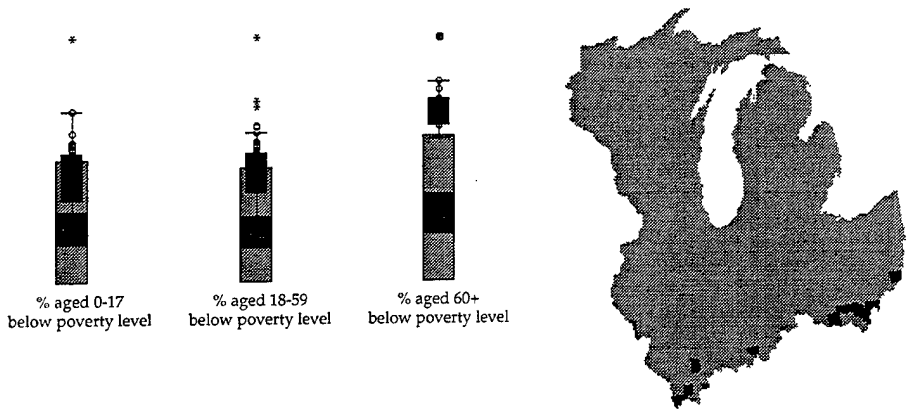


Figure 3: On the left hand is a polygon map of the counties in Bavaria; on the right hand is the trace plot for four variables on kind of land usage.

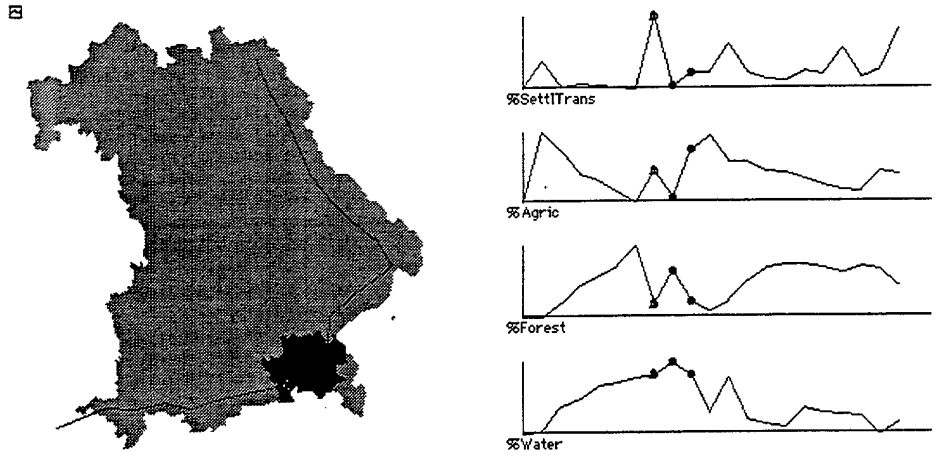


Figure 4: On the left hand is a polygon map of the counties in Bavaria; on the right hand is the trace plot for four variables on kind of land usage.

case of the displayed variable with the corresponding value of the weight variable.

Weighted versions are available for histograms, bar charts and mosaic plots. The plotted areas do not reflect the number of counts in each class, as would be done by the standard plots of this type, but the areas reflect the amount of another variable measured at the objects that fall in a class, see Figure 5. In many surveys such weighted plots help to adjust results and to avoid false impressions that are mainly caused by a specific structure in the underlying sample space.

CONCLUSION

Linking low-dimensional views is an easy interpretable approach to analyze multivariate data. This approach can be straightforwardly extended to spatial data by adding a map to the plot toolbox. Highly interactive environments give full support to the human being who is still the best pattern recognizer. In contrast to linkages between GIS and statistical software, the development of integrated software is easier to use, it offers are more seamless transition from non-spatial to spatial problems, and - most important - it is much faster. In addition, low-dimensional views, like bar charts and histograms, are also familiar to researchers and analysts who are not specialized in statistics. Therefore, communication is much easier to clients when using these types of plots. Further information on the MANET-Project can be found on the World Wide Web under URL

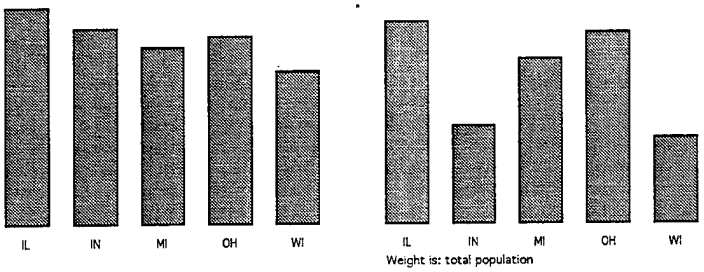


Figure 5: On the left hand we show a standard bar chart where the area of the bars represents number of counties in certain states in the Midwest, whereas on the right hand the area reflects the total population in those counties.

http://www1.math.uni-augsburg.de/ manet/

Future work will concentrate on extensions of the linking paradigm to different types of spatial location. Various data matrices have then to be connected together. In the hierarchical case the procedure is straightforward. Whenever a region is highlighted that is higher in the hierarchy all areas that are included in that region are highlighted too. If an area is highlighted the region containing it is partially highlighted either by using grey scale or by partitioning the region.

ACKNOWLEDGMENT

I am gratefully indebted to the main developers of MANET Antony Unwin, Martin Theus and Heike Hofmann. The work has been completed while the author was visiting the Center for Computational Statistics at George Mason University. I wish to thank Edward J. Wegman and his colleagues at the Center for their kind hospitality and their help. Part of the research on this paper has been supported by DFG grant Wi 1584/1-1 and NSF grant 5-23136.

REFERENCES

1. Openshaw, S. "Developing appropriate spatial analysis methods for GIS." In Maguire D., Goodchild M.F., and D. Rhind, (eds.) Geographical Information Systems: principles and applications, Vol 1. London: Longman, pp. 389-402, 1991.

2. Tukey, J.W. Exploratory Data Analysis. Reading, MA: Addison-Wesley, 1977.

3. Cook, D., Majure, J.J., Symanzik, J., and N. Cressie. "Dynamic Graphics in a GIS: Exploring and Analyzing Multivariate Spatial Data Using Linked Software." Computational Statistics, 11, pp. 467–480, 1996.

4. Bao, S. and L. Anselin. "Exploratory Spatial Data Analysis: Linking SpaceStat and ArcView." In Fischer, M. and A. Getis (eds.), Recent Developments in Spatial Analysis. Berlin: Springer-Verlag, forthcoming.

5. Dykes, J. "cdv: A flexible approach to ESDA with free demonstration software." Manuscript, 1997.

6. Unwin, A. "REGARDing Geographic Data." In Computational Statistics. Heidelberg: Physica, pp. 315–326, 1994.

7. Haslett, J., Wills, G. and A. Unwin. "SPIDER – an interactive statistical tool for the analysis of spatially distributed data." International Journal of Geographical Information Systems, 4, pp. 285-296, 1990.

8. Unwin, A., Hawkins, G., Hofmann, H., and B. Siegl. "Interactive graphics for data sets with missing values – MANET." Journal of Computational and Graphical Statistics, 5, pp. 113–122, 1996.

9. Eick, S. and G. Wills." High interaction graphics." European Journal of Operational Research, pp. 445-459, 1995.

10. Becker, R., Cleveland, W.S., and M. Shyu. "The visual control and design of Trellis Display." Journal of Computational and Graphical Statistics, 5, pp. 123–155, 1996.

11. Monmonier, M."Geographic brushing: enhancing exploratory analysis of the scatterplot matrix." Geographical Analysis, 21, pp. 81–84, 1989.

12. MacDougall, E.B. "A prototype interface for exploratory analysis of geographic data." Proceedings of the Eleventh Annual ESRI User Conference, Vol. 2, pp.547-553 ..

13. Theus, M., Hofmann, H., and A. Wilhelm. "Selection sequences - interactive analysis of massive data sets." In D. Scott (ed.) Computing Science and Statistics, Proceedings of the 29th Symposium on the Interface. Fairfax, VA: Interface Foundation of North America, forthcoming.

14. Craig, P., Haslett, J., Unwin, A., and G. Wills. "Moving Statistics - An Extension of Brushing for Spatial Data." In Computing Science and Statistics, Proceedings of the 21st Symposium on the Interface, pp. 170-174, 1989.

FUZZY CONTROL OF ROBOT COMMUNICATIONS

Aivars Celmiņš
U.S. Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5067
celms@arl.mil

ABSTRACT

This paper deals with the control of radio communications in a battlefield – in particular, of communications among scouting robots. It is assumed that the robots operate in groups and exchange over a common radio channel coded messages about their states, positions, tasks, and observations. Control of the communications is needed to reduce message collisions, and consists of a control of access to the channel. To reduce vulnerability the control should be distributed among the robots such that each robot independently controls its access in a manner that enhances the overall information throughput. Inputs for such a control are approximate statistics of the network status that are obtained by passive monitoring of the message traffic. The control rules are heuristic because the controlled process does not have a set point. These circumstances suggest the use of fuzzy-logic control procedures. The author has developed such control procedures and tested their behavior on a computer model of battlefield communications.

INTRODUCTION

A typical communications network in a battlefield consists of a moderate number of nodes that are broadcasting on a single low-bandwidth radio channel. Problems arise when more than one node tries to access the channel at the same time. The ensuing message collisions can cause a breakdown of communications just at those times when information exchange is important.

To ensure smooth communication and to enhance the information throughput rate, a control of network access is mandatory. However, a hierarchical control is not practical in a battlefield environment for two reasons. First, the status of the network changes dynamically. Therefore, a regulating node must be constantly supplied with information about the present status. Such information gathering would use up broadcasting time, and the received information might be outdated on arrival. Second, the concentration of control in a single node makes the communications network more vulnerable.

These difficulties can be avoided by using a distributed control where each node listens to the network traffic and makes independent decisions for accessing the network. This is possible because in a modern battlefield communication system, the nodes are computers and the communication is in digital form among these computers. That is, each node has ample computing power to analyze the network traffic and make decisions.

The inputs for the control procedures are obtained by passive listening that provides only approximate, albeit up-to-date, information about the network status. Therefore, the control rules must be of such type that allows approximate input. Also, the goal of the control is defined only approximately as "enhance the information throughput." The system has no set point, and control rules that are based on error terms are not applicable. This suggests the use of fuzzy-logic control rules. This paper describes such control procedures that have been developed at the U.S. Army Research Laboratory.

Approved for public release; distribution is unlimited.

NETWORK ACCESS ALGORITHM

We are seeking access control procedures with the following properties:

1) All nodes have equal chances for access.
2) Information throughput is enhanced.
3) Priority messages have higher access chances.

An algorithm that achieves these goals can be constructed around the following broadcast procedure. A monitor at each node constantly listens to the radio channel and informs the access manager program when the channel is free. At that time, t, the manager determines which message from its message queue should be broadcast, chooses a random time interval, Δ, from a prescribed access delay interval range (0,D), and sets a tentative broadcast time, $t + \Delta$, for the message. If the channel is still free at the set time, then the message is broadcast. Otherwise, the manager aborts the procedure and waits for the next free time signal.

Obviously, the procedure satisfies the first property if all nodes use the same access delay interval, D. The procedure would also prevent message collisions, if the response time of the nodes would be zero and the propagation speed of radio signals infinite. In practice, however, the response times are finite, and it is estimated that messages from different nodes will be broadcast simultaneously and collide if their set times $t + \Delta$, differ by less than 0.5 s. (A typical message length is of the order of 1 to 10 s.) Collisions reduce the information throughput rate because collided messages must be repeated. The probability of collisions is reduced by increasing the size of the access delay interval D. On the other hand, a large D means long idle times for the network and the idle time is reduced by reducing D. Hence, there exists an optimal D that corresponds to a maximum throughput rate and somehow depends on the size and state of the network. The goal of the control procedure is to find, in real time, an optimal D from information that can be obtained by monitoring the network traffic.

The third property is taken care of by reducing the global D for high-priority messages and increasing it for low-priority messages. The priorities are assigned automatically by the message-generating programs. (Typical information exchanged by scouting robots consists of data about the robots' positions, states, and current tasks; descriptions of observed targets; and changes of environments in the vicinities of the robots.)

INPUT INFORMATION

The state of a battlefield communications network is completely described by the number and state of the nodes, the message queue at each node, and the characteristics of external noise. However, the current values of these state parameters are not available to the access managers at the nodes. Instead, the network conditions must be inferred from approximate information that is obtained by listening to the network traffic. We have chosen three groups of network parameters that can be obtained by listening. These parameters represent time averages of observations during a listening time interval, L, prior to the reference time t. One group consists of the relative usage time of the radio channel during the time interval $(t - L, t)$ in three usage categories. A second group consists of the relative number of network accesses during the same time interval in the three categories. A third group consists of average message lengths in the three categories. The categories of network usage time are:

1. Idle time.
2. Successful transmissions.
3. Collided transmissions.

For effective control, one also needs predictions about the expected status of the network. Such predictions can be obtained from the trends of the observed parameters that usually are obtained by numerical differentiation. However, in the present problem, simple numerical differentiation cannot be used because of the oscillatory character of the data. Therefore, trend indicators were obtained by subtracting from the simple averages of the parameters corresponding weighted averages over the same averaging interval L with linearly decreasing weights.

The length L of the listening interval cannot be chosen arbitrarily. An excessively short length provides only useless data about the instant channel status, while an averaging over an excessively large L is not sufficiently responsive to changing trends of the averaged parameters. Therefore, the access controller also controls the listening and averaging interval L by adjusting it as necessary if network conditions change.

CONTROL RULES

The network access algorithm is controlled by the length D of the access delay interval. The choice of D is in turn influenced by the length L of the monitoring and averaging interval. The procedures for the control of these two access parameters were formulated in terms of fuzzy-logic rules. In this section, we provide a short description of these rules. A more detailed description and discussion is found in reference [2].

CONTROL OF MONITORING

The proper size of the monitoring and averaging interval L was found by experiments to be about 40 to 50 times the average length of transmitted messages. (These and other experiments were done with a computer model of battlefield networks [1].) To accommodate changing network conditions, L was controlled dynamically; that is, L was increased or decreased from its current value if it deviated significantly from the set value of 50 times the current message length. The deviation was expressed in terms of the dimensionless quotient

$$Q = L / (50 \cdot a) - 1 \ , \tag{1}$$

where a is the average message length. The correction of L_{old} was expressed by a factor λ as follows:

$$L_{new} = L_{old} \cdot (1 + \lambda) \ . \tag{2}$$

The control rules for the computation of the corrector λ for given deviation Q are summarized in the following fuzzy-rule table.

Q	NL	N	Z	P	PL
λ	PL	P	Z	N	NL

Here, NL, N, Z, P, and PL denote fuzzy sets that define the categories "negative large," "negative," "zero," "positive," and "positive large," respectively. The first rule in the rule table is

If Q is "negative large," then make λ "positive large."

The other four table entries denote corresponding rules. The membership functions of the fuzzy sets that describe the categories were determined by numerical experiments with the battlefield network model.

CONTROL OF ACCESS

The access to the radio channel is regulated by the size of the access delay interval D. In principle, the size of D can be efficiently controlled by the simple algorithm

$$D_{new} = D_{old} \cdot (1 + \delta) \ , \tag{3}$$

where δ is an output of the control rules. In practice, this simple procedure must be modified to ensure that after an initialization time, all independent controllers indeed produce the same value of D, even when starting from different initial conditions. For clarity, we shall first describe the control rules in the context of Eq. (3) and discuss the modifications of the algorithm later.

Let $c - i$ be the difference between the relative uses of channel time for colliding messages and for idling, respectively, during the listening interval L. A common-sense rule for the governing of D is "increase D if

$c - i > 0$ and decrease D if $c - i < 0$." The rule table is as follows.

$c - i$	NL	N	Z	P	PL
δ	NL	N	Z	P	PL

This set of rules turns out to be very effective, but it does not account for the possibly different message lengths. Thus, a single collision involving a very long message can increase the average collision time, c, as much as many collisions involving short messages, but in each of theses cases, the proper control strategy is different. Therefore, as a second input, the trend of the number of colliding accesses was used. Let τ_c be the observed trend. Then the rule table is as follows.

τ_c	NL	N	Z	P	PL
δ	NL	N	Z	P	PL

To make the control more responsive to extreme conditions, two more inputs were considered: "idle time over a threshold" and "collision time over a threshold." The rule table for the former was as follows.

$i - t_i$	NL	N	Z	P	PL
δ	O	O	O	NL	NL

Here, O indicates "no output"; that is, no rule is fired in these cases. These rules provide an acceleration of the reduction of D if i is unreasonably large. The corresponding rule table for the collision time over a threshold is as follows.

$c - t_c$	NL	N	Z	P	PL
δ	O	O	O	PL	PL

Experiments show, however, that the last two sets of rules have only a minor effect on the performance of the control if used in addition to the first two sets of rules. If used without the first two sets of rules, then the performance of the control was not as good as with the first two sets alone.

The outputs from the rules were combined using the compositional rule of inference [3 - 6], and the fuzzy result was defuzzified using the center of gravity method [5]. A combination of rules in the form of two-dimensional rule tables was also tried and found to perform equally well.

The described control works efficiently if all nodes start with the same initial value of D and have the same observational input. In practice, one can only assume that the monitored inputs are approximately equal for all nodes, but the initial conditions can be quite different because different nodes might join the network at different times. Therefore, the updating formula (3) for D must be modified such that after an operation over a reasonable time (several minutes), differences among the nodes become negligible. One method to achieve this is to replace Eq. (3) by an absolute output instead of the corrective one. However, experiments indicate that such a control is not efficient because absolute outputs cannot be easily adapted to changing network conditions. We now describe a replacement of Eq. (3) that was found satisfactory.

First, we separate in Eq. (3) the dimensional factor D from the nondimensional corrector $F = 1 + \delta$, and devise for F an updating procedure such that the factor drifts with time to a fixed value that is independent of its initial value. Let Δt [s] be the difference between the current and previous times of network parameter readings. (Network parameters are updated at discrete times after the end of each activity – that is, at the end of each transmitted message – because only at those instants the average usage times and access numbers can be computed.) Let f be a fixed value of F to which the correction factor should drift with increasing time. Let $\varepsilon = \exp(-\Delta t/60)$. Then the factor is updated as follows:

$$F_{new} = [\, f \cdot (1 - \varepsilon) + F_{old} \cdot \varepsilon \,] \cdot (1 + \delta) \ . \tag{4}$$

One can show that the exponential factor in the formula has the effect that the starting value of F becomes insignificant after a few minutes. Hence, if all nodes would use the same f, then equalization among the nodes would result. Experiments show, however, that an a priori assignment of a unique value of f for all conditions is not desirable. Rather, to have efficient access control, its value should be allowed to vary typically between 0.001 and about 80. We achieve equalization of f by assigning to it only a limited number of discrete values within this range. In particular, the access controllers assign to f values in steps of four according to the following algorithm:

If $\delta_{old} > 0$ and $\delta_{new} > 0$, then increase f;
If $\delta_{old} < 0$ and $\delta_{new} < 0$, then decrease f.

Equalization of the factor f among the nodes occurs in short time because typically the control output δ continuously increases or decreases over more than 20 readings.

A reasonable value for the dimensional factor D_{old} is, for instance, the average idle time. However, that average increases when D is increased and vice versa, thereby accelerating any corrections by the control. In extreme conditions, this results in a drift of D to zero or infinity. We avoid such excursions by using the logarithm of the average idle time instead of the average itself. The average idle time is also subject to smaller oscillations between readings that are not conducive to access control. These oscillations were reduced by using a weighted average of the current and previous reading of the average idle time. The weighting was again done with an exponential factor that reduces the influence of old readings. Let $\upsilon = \exp(-\Delta t/300)$. Then the effective value of the average idle time is computed by

$$i_{eff} = (i_{old} \cdot \upsilon + i_{new}) \ / \ (\upsilon + 1) \ . \tag{5}$$

The final formula for the access delay time that replaces Eq. (3) is with these modifications

$$D_{new} = [\, 0.01 + \log(1 + i_{eff}/4) \,] \cdot F_{new} \ . \tag{6}$$

GRANULATION

In all sets of rules, five categories of input and output were found to be sufficient. The membership functions that define the categories were assumed to be trapezoids. The utility programs that were developed and used for rule combinations and defuzzifications accept arbitrary forms of membership functions, but it was found that more general functions are not needed to achieve optimal control. Numerous experiments were carried out to determine optimal definitions of input/output categories under different conditions. Results of these experiments indicate that the best membership functions are robust in the sense that similar functions perform well for a wide variety of networks and that the corresponding minima are flat. If the control should be implemented in field equipment, then special fuzzy-logic chips would be used for the fuzzy-logic operations. The results of the numerical experiments will provide guidance for the design of such chips.

Figure 1 shows the optimal categories for the rules involving the difference $c - i$ and the trend τ_c, and Figure 2 shows the output categories of the control parameter δ. It is interesting to note slight asymmetries in the granulations of the inputs. These asymmetries were not chosen arbitrarily but developed by the tuning process of the granulations. One of the effects of the asymmetries is that, in general, the collision time is kept at a fraction of the idle time.

EXAMPLE

Let the network consist of n nodes, and let us assume that the message generation rate is sufficiently high so that all nodes accumulate queues of messages for broadcasting. Further let the message generation stop at a preset time. Then a good measure for the performance of access control is the time that is needed to empty all queues. Using different control strategies with the same set of message queues, one can compare the control

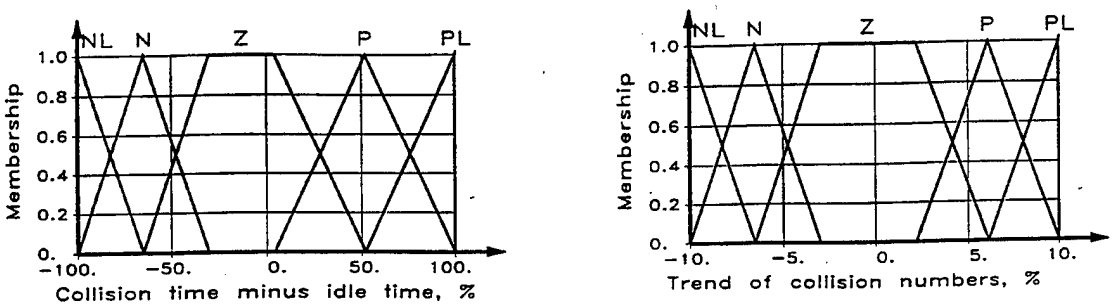


Figure 1. Input categories.

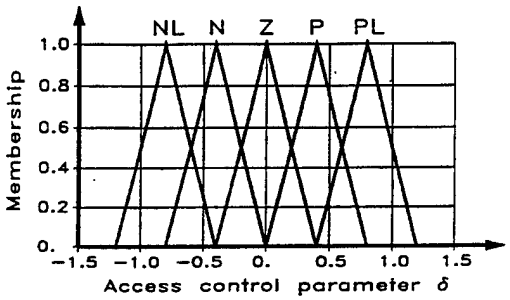


Figure 2. Granulation of control output.

performances by comparing the corresponding completion times. We present an example of a network with four nodes. The information content, measured by the total length of all messages in a queue, was approximately equal for all nodes, but the average lengths of messages were different for different nodes. In Figure 3, the results of an experiment with fixed access delay time are compared with the results with controlled delay time. The left-hand graph shows curves of network usage times for a fixed access delay time D of 13 seconds. All queues are emptied in about 34 minutes. The graph shows that during that time the channel has been used for message transmissions about 17 minutes (dashed curve), for idling about 11 minutes (solid curve), and for colliding messages about 6 minutes (dotted curve). It is obvious that during the first 20 minutes the access delay time interval D was too short, causing many message collisions. On the other hand, during the last 10 minutes of the experiment, there were no new collisions (the dotted curve is flat), and a smaller D could be used to reduce the idle time. The right-hand graph shows the same network with a controlled delay time interval D. We observe that the dynamically adjusted D equalizes the collision and idle times and achieves a shorter completion time. In this experiment, the dynamically adjusted access delay time D was varied between less than 1 s and about 65 s.

CONCLUSION

The distributed control of battlefield communications has been found efficient, and the optimal parameters of the investigated control algorithm were found to vary little under very different conditions. This result makes the design of an efficient general fuzzy-logic control of battlefield communications possible.

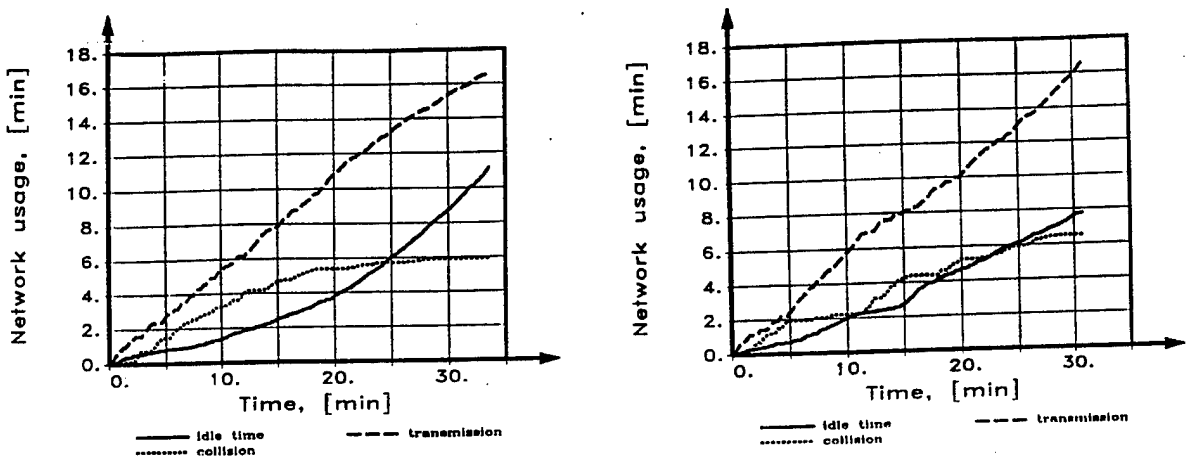


Figure 3. Network usage with fixed and controlled access.

REFERENCES

[1] Celmiņš, A. Battlefield Communications Network Model BATNET, U.S. Army Research Laboratory Memorandum Report ARL-MR-244, August 1995.

[2] Celmiņš, A. Fuzzy-Logic Control of Battlefield Communications, U.S. Army Research Laboratory Technical Report ARL-TR-1215, October 1996.

[3] Klir, G. J. and B. Yuan. Fuzzy Sets and Fuzzy Logic, Prentice Hall PTR, Upper Saddle River, NJ, 1995.

[4] Pedrycz, W. Fuzzy Control and Fuzzy Systems, 2nd, Extended Edition, John Wiley & Sons, New York, NY, 1992.

[5] Terano, T., K. Asai, and M. Sugeno. Fuzzy Systems Theory and its Applications, Academic Press, San Diego, CA, 1991.

[6] Zadeh, L. A. "The Concept of a Linguistic Variable and its Application to Approximate Reasoning, Part III," Information Sciences, vol. 9, p. 43, 1975.

REAL-TIME ANIMATION OF PARTICLE BEHAVIORS

Jim X. Chen and Edward J. Wegman[1]
Department of Computer Science and Center for Computational Statistics
George Mason University

ABSTRACT

Real-time simulation of particle behaviors is very useful in training, education, art, advertising, and entertainment. There is no successful model for realistic dust behaviors generated by a traveling vehicle. In this paper, we use particle systems and behavioral simulation techniques to simulate these dust behaviors in real-time. First we analyze the forces and factors which affect the dust generation and the behaviors after dust particles are generated. Then we construct physically-based empirical models to generate dust particles and control the behaviors accordingly. After that, we further simplify the numerical calculations by dividing the dust behaviors into three stages, and establishing simplified particle system models for each stages. We employ motion blur, particle blending, texture mapping, and other computer graphics techniques to achieve the final results. Our major contribution includes analyzing dust behaviors in detail, constructing physically-based empirical models that correlate the behaviors to the dust generating forces and other factors, and that achieve simulations in real-time.

INTRODUCTION

In many virtual environments and distributed interactive simulations, it is desirable to simulate trucks, tanks, armored vehicles, bulldozers, and other ground moving objects. However, typically dust behaviors are not generated when these objects travel on an unpaved road. Dust behaviors caused by different factors (such as natural wind and a fast traveling vehicle) appear everywhere. Simulating physically realistic, complex dust behaviors is very useful in interactive graphics applications, such as computer art, advertising, education, entertainment, and training. However, due to the lack of modeling and simulation techniques and methodologies, there is currently no successful real-time simulation for realistic dust behaviors. As computers and their graphics systems become much faster and more powerful, many natural phenomena (such as the behaviors of fluids, terrains, trees, fireworks, volcanos, clouds, etc.) are simulated in real-time [1, 6, 5, 7, 8, 10, 11, 12, 13]. We believe it is appropriate now to include dust behaviors into real-time simulation.

Hsu and Wong [3] introduced a dust accumulation model. Their model presents static appearance of dust accumulation without behavior and animation. Cowherd, Williams, and other researchers [2, 14] studied dust and the mechanisms of dust generation. Their purpose was to study and measure the density of the dust in the real battlefield instead of simulating the dust behaviors in graphics. Today, military training in graphics and distributed interactive simulation is one of the major topics for research and applications [4], and generating dust behaviors in real-time significantly increases the realism of the simulated training environment.

In this paper, we introduce a method for simulating the dust behaviors caused by a fast traveling vehicle in real-time. The method is a combination of particle systems and behavioral simulation techniques. The *Particle systems* technique was first introduced to computer graphics by Reeves [8] in 1983, and is now widely used to simulate fuzzy or dynamic objects, such as fire, grass, explosions, clouds, water, trees, etc. These objects have no fixed shape and sometimes change their shapes or behaviors stochastically. They have ill-defined boundaries that make surface-based modeling impractical. It is apparent that dust behaviors behind a moving vehicle belong to this category. The *Behavioral simulation* technique uses a physically-based modeling method to calculate and update the object's state, and draw the object repetitively after each calculations to achieve the behavior animation in real-time. We also employ motion blur for small and fast moving particles, particle blending instead of hidden-surface removal, texture mapping, and other graphics techniques to achieve better performance and appearance of the final results.

---

1. Approved for public release; distribution is unlimited

The work of Dr. Wegman was supported by the Army Research Office under contract DAAH04-94-G-0267 and by a National Science Foundation Group Infrastructure Grant DMS-9631351. This work was completed while Dr. Wegman was a Navy-ASEE Distinguished Faculty Fellow at the Naval Surface Warfare Center, Dahlgren.

In order to build up a physically-based realistic simulation, we first analyze the forces and factors which affect the dust generation and the behaviors after dust particles are generated, and then construct physically-based empirical models to generate dust particles and control the behaviors accordingly. However, the models are time-consuming and inefficient. Therefore, based on the models and analysis of the forces, we further simplify the numerical calculations by dividing the dust behaviors into three stages, and establishing simplified particle system models for each stages. The resulting models are satisfactory for real-time simulation as well as achieving realistic dust behaviors. Our major contribution includes analyzing dust behaviors in detail, constructing physically-based empirical models that correlate the behaviors to the dust generating forces and other factors, and that achieve simulations in real-time. Our work is a useful addition to many applications in simulated virtual environments, including military simulation and training.

DUST BEHAVIORS

In this section, we first discuss and analyze how dust particles are generated; then we introduce the factors which affect the dust behaviors, after that we analyze the forces acting on a dust particle and establish corresponding physically-based empirical models to calculate and update the dust behaviors.

DUST GENERATION

As a vehicle wheel passes over an unpaved surface, three basic forces are developed - vertical pressure, horizontal stress, and friction. Vertical pressure, which is due to the weight of the vehicle ($\boldsymbol{WT}_{car}$), will produce ground surface vibration and/or deformation, crushing large particles into smaller ones. Horizontal stress and friction, which are largely due to the driving power which sustains the velocity ($\boldsymbol{V}_{car}$) and acceleration of the vehicle, will further comminute the particles and carry them on the surface of the tire. The slippage between the tire and the ground surface will lift particles of different sizes due to the adhesive and shear forces, and eject them at different places on the tire surface due to the centrifugal forces.

The ground vibration and/or deformation will also eject fine particles. The dust particles are then either entrained in the turbulent air behind the vehicle or return to the ground depending on their properties and conditions. Bigger particles will fall back to the ground surface more rapidly while the fine ones will remain suspended in the air drifting with the current. Small stones and blocks of muds will fall back to the ground immediately after ejection from the tires, and will bounce up and down, also generating dust into the air.

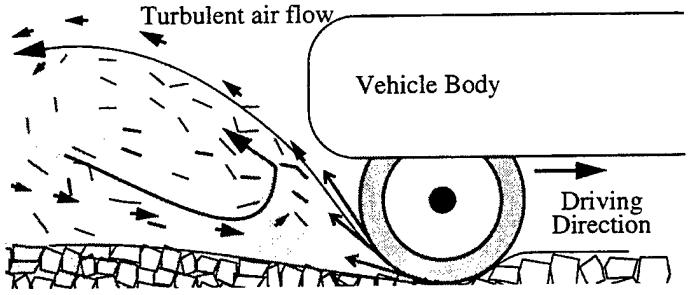


Figure 1: Dust generation process

There are many other important factors that affect dust generation, such as the material and composition of the ground surface (Mat_p), the size and properties of the vehicle (S_{car}), and the conditions of the environment ($\boldsymbol{F}_{env}$).

The material and composition of the ground surface (Mat_p) depend on the density (D_p) and moisture (MO_p) of the ground surface, the average size (S_p), mass, and adhesion of each individual dust particle. If the ground is wet and the average size of the particles is large, then there will be fewer particles. If the dust density of the ground surface is high, then there will be more particles. We use the following equation to measure this parameter:

$$Mat_p = \frac{D_p}{\alpha_1 \cdot S_p + \alpha_2 \cdot MO_p} \tag{1}$$

where $\sum_{i=1}^{2} \alpha_i = 1$, and α_i is the weight coefficient of the corresponding parameter, for $i = 1, 2$. The size and properties of the vehicle (S_{car}) depend on the weight ($\boldsymbol{WT}_{car}$), height (H_{car}), and width (W_{car}) of the vehicle. Heavier and bigger vehicle will generate more dust particles. We use the following equation:

$$S_{car} = \beta_1 \cdot |\boldsymbol{WT}|_{car} + \beta_2 \cdot H_{car} + \beta_3 \cdot W_{car} \tag{2}$$

where $\sum_{i=1}^{3} \beta_i = 1$, and β_i is the weight coefficient of the corresponding parameter, for $i = 1, 2, 3$. The conditions of the environment ($\boldsymbol{F}_{env}$) including humidity, air pressure, and many other environment damping forces will also affect the dust generation.

In summary, most dust particles are generated right behind the wheels. Some fine dust particles are lifted from the ground surface because of the turbulent wake behind the vehicle. Each particle is generated with its own initial mass, size, and velocity. The number of dust particles generated by the vehicle for each simulating frame is calculated as follows:

$$N_p = \frac{|\boldsymbol{V}_{car}|(\gamma_1 \cdot Mat_p + \gamma_2 \cdot S_{car})}{|\boldsymbol{F}_{env}|} \tag{3}$$

where $\sum_{i=1}^{2} \gamma_i = 1$, and γ_i is the weight coefficient of the corresponding parameter, for $i = 1, 2$. Once the particles are entrained in the turbulent wakes behind the vehicle, their behaviors are affected by similar factors which will be discuss in the next section.

FACTORS AFFECTING DUST BEHAVIORS

There are numerous factors that affect the dust behaviors caused by a fast traveling vehicle. Here we summarize the important factors that have more serious effects on the dust behaviors.

When a vehicle travels quickly, it produces a 3D volume behind where the atmosphere pressure is lower than that of the other areas. The turbulent wake is mostly inside this volume and generates vigorous dust behaviors. The shape and size of the 3D volume are mainly decided by the velocity, height, and width of the vehicle. At the same time, the properties of each individual dust particle will affect its own behaviors. These properties include the shape, size, mass, and initial conditions. Intuitively, a particle will fall back to the ground faster if it is heavier and smaller. In addition, the humidities, wind, and environmental damping forces have an effect on the dust behaviors also. We summarize the parameters which have the greatest influence on the dust behaviors in table 1

Table 1: Parameters affecting the dust behaviors

Items	Parameters	Description
Vehicle	Velocity — $\boldsymbol{V}_{car}$ Height — $H_{car;}$ Width — W_{car}	Decide the size of the 3D volume behind a car where the turbulent wake is generated.
Dust particle	Size — $S_{p;}$ Mass — m_p	Affect how an individual particle will react to the external forces
Environment	Wind — $\boldsymbol{V}_{air;}$ Moisture — MO_{air}	Influence the dust behaviors in general.

DYNAMICS OF A DUST PARTICLE

Computational fluid dynamics could be used to calculate the exact turbulent wake behind the vehicle. However, this approach is computationally complex and prohibits achieving simulation in real-time. In order to describe the dust behaviors caused by a fast traveling vehicle, here we simplify the dynamics of a dust particle, analyze all the important forces, correlate these forces with the parameters affecting the dust behaviors, and establish analytical models which can be used to simulate the dust behaviors in real-time. Here we first analyze the forces behind a traveling vehicle, then we study the effects of these forces on an individual particle.

As the vehicle travels forward quickly, it produces a 3D volume behind the vehicle where the atmosphere pressure is smaller than that of the other areas. To simplify the situation for our analysis, let's assume that the vehicle does not turn, and the area affected by the vehicle is a box. That is, our particle systems' range is a box moving at the speed of the vehicle. The box's height and width are the same as those of the vehicle's (H_{car}, W_{car}), and its length (L) depends on the vehicle's velocity. Because of the fast movement of the vehicle, different places within the box have different atmosphere pressures. The differences among the pressures will generate turbulent wakes, and the dust behaviors accordingly (Fig. 2).

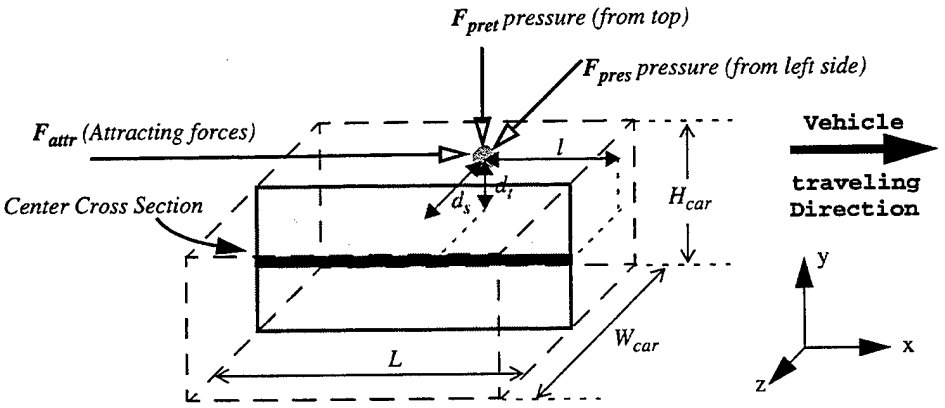


Figure 2: The pressures (forces) generated

There are five kinds of forces acting on any particle inside the box area:

- The pressure or attraction force F_{attr} towards the rear side of the vehicle
- The atmosphere pressure force F_{pres} from the two sides of the box area
- The atmosphere pressure force F_{pret} from the top and bottom of the box area
- The atmosphere damping force F_{air} against the particles relative movement
- The dust particle's gravity F_{grv}

F_{attr} is a function of the vehicle's velocity (V_{car}), the size of the vehicle (S_{car}), and the distance between the dust particle and the rear side of the vehicle (l). As the vehicle moves ahead, l becomes larger, and therefore F_{attr} is reduced rapidly. Its direction is approximated by the direction which the vehicle travels. We have the following equation:

$$F_{attr} = \zeta\left(\frac{V_{car} \cdot S_{car}}{l}\right) \tag{4}$$

where ζ is a constant scale parameter. F_{pres} is a function of the vehicle's velocity (V_{car}), the distance between the dust particle and the rear side of the vehicle (l), and the distance between the particle and the horizontal cross section in the center of the box (d_s). As the vehicle moves ahead, l becomes larger, and therefore F_{pres} is reduced rapidly. Its direction is towards the center cross section and parallel to the ground, which is approximated by the direction of d_s. We have the following equation:

$$F_{pres} = \vartheta_1\left(\frac{|V_{car}| \cdot d_s}{l}\right) \tag{5}$$

where ϑ_1 is a constant scale parameter. F_{pret} is similar to F_{pres}. It is a function of the vehicle's velocity (V_{car}), the distance between the dust particle and the rear side of the vehicle (l), and the distance between the particle and the horizontal cross section in the center of the box (d_t). Its direction is towards the center cross section and perpendicular to the ground, which is approximated by the direction of d_t. We have the following equation:

$$F_{pret} = \vartheta_2\left(\frac{|V_{car}| \cdot d_t}{l}\right) \tag{6}$$

where ϑ_2 is a constant scale parameter. F_{air} is a function of the particle's velocity (V_p), the particle's size (S_p), the environment wind (V_{air}), and the air moisture (MO_{air}). Its direction is against the particle's movement in opposite to V_p. We have the following equation:

$$F_{air} = (\delta_1 \cdot V_{air} - \delta_2 \cdot V_p)(\eta_1 \cdot S_p + \eta_2 \cdot MO_{air}) \tag{7}$$

where δ_1, δ_2, η_1, and η_2 are constant parameters.

In summary, all the forces acting on a particle are shown in Fig. 3. Here we ignore the collisions among the dust particles.

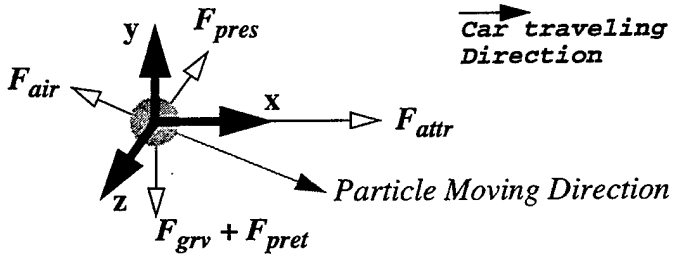


Figure 3: Forces acting on a dust particle

Let F_p be the force acting on a dust particle, P the position, V_p the velocity, A_p the acceleration, and m_p the mass of the particle. Then a dust particle's behavior is described by the following equations:

$$F_p = F_{attr} + F_{pres} + F_{pret} + F_{grv} + F_{air} \tag{8}$$

$$A_p = \frac{F_p}{m_p} \tag{9}$$

$$V_p = V_0 + \int_{t_0}^{t_1} A_p dt \tag{10}$$

$$P = P_0 + \int_{t_0}^{t_1} V_p dt \tag{11}$$

To simplify the calculation, we use Euler's method to approximate the particle's next state:

$$V_i = V_{i-1} + A_p \cdot \Delta t \tag{12}$$

$$P_i = P_{i-1} + V_i \cdot \Delta t \tag{13}$$

The algorithm to compute the solution to the dust behaviors then is as follows. For the known current state of a particle $\{V_{i-1}, P_{i-1}\}$, the next state $\{V_{i-1}, P_{i-1}\}$, after Δt time, is calculated by Equation (12) and (13). These equations use functions (4) to (9). Equation (1) to (3) are used to generate a number of dust particles. We have frames of the dust behaviors as in Fig. 4. Changing the parameters and conditions, we can achieve different behaviors and appearances to suit the needs of the applications. The simulation is at about 8 frames per second.

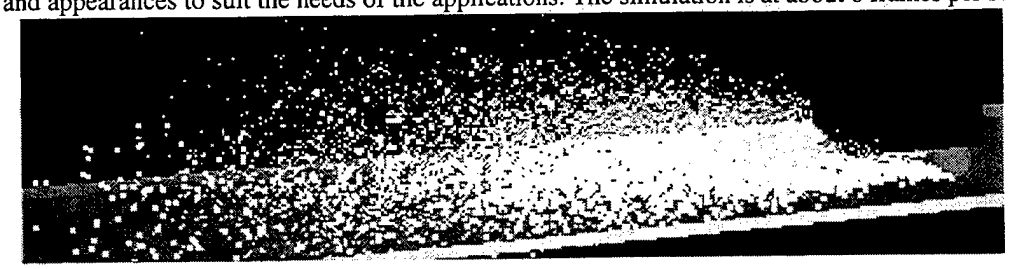


Figure 4: A frame of dust simulation

SIMPLIFIED DUST PARTICLE SYSTEMS

The above physically-based dust particle model is time-consuming for a large number of particles. There are too many factors in the equations, and the forces on each individual particle have to be calculated during the whole simulation period. There are redundant calculations because when the particles are further away from the vehicle, the forces F_{attr}, F_{pres}, and F_{pret} are all reduced to near zero. Based on the above analysis (equation 4-7), we can divide a dust particle's behaviors into three stages to further simplify the simulation. We consider that we have three different particle systems (models) working together to simulate the dust behaviors. The three stages are called *turbulent vortex, inertial momentum*, and *airborne drift*, respectively.

TURBULENT VORTEX (THE FIRST STAGE)

Once a particle is generated, the initial forces F_{attr}, F_{pres}, and F_{pret} acting on it are very large (see Fig. 2, Fig. 3, and equation 4-7). F_{attr} causes the particle to move in the forward direction, F_{pret} causes the particle to move up and down, and F_{pres} causes the particle to move left and right. All other forces are relatively small at this time. Suppose the particle is located in side 1 of the box area, because the particle has a side pressure force F_{pres} pointing towards side 2, the particle will move from side 1 to side 2. Once the dust particle goes across the center section into side 2, F_{pres} will change its direction, and the particle will accelerate and move back from side 2 to side 1. It is similar for F_{pret} but in a perpendicular direction. At this stage, the forces acting on the dust particles are relatively strong. Overall, depending on the initial velocity, the particle will behave as in a turbulent vortex shown in Fig. 5.

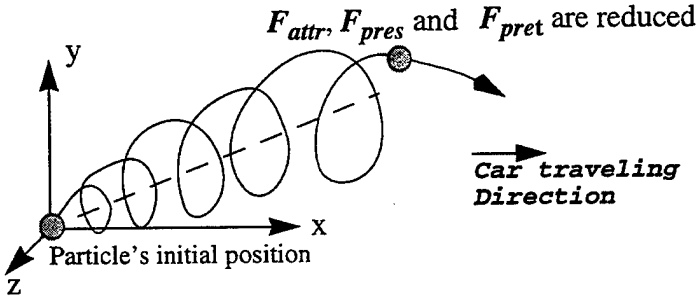


Figure 5: Dust particle traveling trace (the first step)

Instead of calculating all the forces, we can simplify the model by just simulating the turbulent vortex with some random behaviors at this stage. Particles are rotating around the center of the vortex. Here we assume the center of the vortex is the center of the cross section in the box area (Fig. 2). Its height is $H_{car}/2$. The angle of rotation (*RotAngle*) for each frame of simulation can be calculated by the following simplified equation:

$$RotAngle = \frac{RotD}{|\boldsymbol{R}|^{Tightness}} \tag{14}$$

where $\boldsymbol{R} = \boldsymbol{d}_s + \boldsymbol{d}_t$, which is the distance from the particle to the center of rotation (we use bold to represent vectors). Its initial value is a function of N_p (equation 3). We use the following equation to adjust $\boldsymbol{d}$ for each simulation frame:

$$|\boldsymbol{R}|_i = |\boldsymbol{R}|_{i-1} \cdot (1 + \Delta r) \tag{15}$$

$$\Delta r = dR + k_1 \cdot \frac{m_p}{S_p} \cdot myRandom(-\varpi, 0, \varpi) \tag{16}$$

where dR is a constant related to the initial velocity towards the center of the rotation, k_1 is a constant coefficient, m_p is the mass of the particle, S_p is the size of the particle, function *MyRandom*(a, b, c) returns a random pick of the numbers enclosed, and ϖ is a comprehensive weight parameter reflecting the vehicle's properties:

$$\varpi = \frac{\delta_1 \cdot S_{car} + \delta_2 \cdot |V_{car}|}{S_{car} + |V_{car}|} \tag{17}$$

where $\sum_{i=1}^{2} \delta_i = 1$, δ_i is the weight coefficient of the corresponding parameter, for $i = 1, 2$; $0 \le \varpi \le 1$. The value *Tightness* is usually between 1.0 and 2.0. Higher *Tightness* causes *RotAngle* to fade away more quickly when the distance becomes larger. *RotD* is the distance the particle traveled at each time frame around the vortex circle, perpendicular to the center of vortex. We use the follow equation to calculate this value:

$$RotD_i = RotD_{i-1} \cdot (1 + \Delta D) \tag{18}$$

$$\Delta D = dD + k_2 \cdot \frac{m_p}{S_p} \cdot myRandom(-\varpi, 0, \varpi) \tag{19}$$

where dD is a constant related to the initial velocity perpendicular to the center of the rotation, and k_2 is small constant. A dust particle may be below ground as its rotating radius increases. In our simulation, we just remove those particles which hit the ground.

The particle's translation distance (*TransD*) along the vortex axis is calculated as follows:

$$TransD_i = TransD_{i-1} \cdot (1 + \Delta TD) \tag{20}$$

$$\Delta TD = dTD + k_3 \cdot \frac{m_p}{S_p} \cdot myRandom(0, \varpi) \tag{21}$$

where dTD is a constant related to the initial velocity parallel to the center of the rotation, and k_3 is a small constant.

INERTIAL MOMENTUM (THE SECOND STAGE)

As the vehicle travels and time passes by, the forces $\boldsymbol{F}_{attr}$, $\boldsymbol{F}_{pres}$, and $\boldsymbol{F}_{pret}$ reduce rapidly, and finally disappear. At this moment, the particle will continue its movement at its current momentum. The forces $\boldsymbol{F}_{grv}$ and $\boldsymbol{F}_{air}$ are the primary forces governing its behavior (acceleration.) We call this stage the Inertial Momentum stage. This stage will continue until dust particle's velocity is reduced such that the particle's velocity becomes a small constant.

Every dust particle has a lifetime parameter, which is used to decide when its motion enters from the first stage into the second stage, and from the second to the third. We use the following equation to simulate the dust particle movement at this stage. The initial velocity is calculated from the first stage's values.

$$A_i = \Lambda(m_p, S_p, V_{i-1}) \tag{22}$$

$$V_i = V_{i-1} + A_i \cdot \Delta t \tag{23}$$

$$P_i = P_{i-1} + V_i \cdot \Delta t \tag{24}$$

where Λ is a function to calculate the acceleration. Here we use $\Lambda(m, S, V) = \frac{VS}{m}$. When A_i becomes very small, we get into the drifting stage, as discussed in the next section.

AIRBORNE DRIFT (THE THIRD STAGE)

When the total force on a particle becomes very small, the dust particle begins to drift with constant velocity. The forces acting on the particle are balanced. Most of the dust particles will drift with the wind and eventually fall back to the ground. In this stage, the particles stay in the air are those with very small mass quantities. We simply keep the constant velocity with some random disturbances. If a dust particle touches the ground, it is dead. If a dust particle drifts from the range of the particle systems to the outside area, it is considered dead. Dead particles are faded away after a few frames of simulations.

RENDERING TECHNIQUES AND RESULTS

MOTION BLUR

We use motion blur to achieve better animation. We record every dust particle's several continuous positions. Each dust particle has a head pointer which is the current position, and a tail pointer which is the fading position. A particle is drawn a number of times into the buffer with bigger and bigger fading coefficients. The head is drawn at its current position with the particle's original color, and the tail is drawn at the earliest position with a much dimmer color. We let these smaller particles to have more blurring effect. That is, the smaller the dust particle is, the longer blur process will be. This simple technique seems to make the simulation more realistic. We also use the comprehensive parameter ϖ (equation 17) to control the blurring process. Larger ϖ value causes particles having more blurring effects. We can modify different parameter for blur according to the simulation applicable for certain application.

BLENDING

Dust particles can obscure other particles that are behind them, or they can be transparent and can cast shadows on other dust particles. We uses Reeves's method to deal with this situation. Every particle is treated as a point light source when it is displayed. Each particle adds a bit of light to the pixels that it covers. A particle behind another particle is not obscured but rather adds more light to the pixels covered.

In order to speed up the rendering process, we restrict our rendering area to be the box area shown in Fig. 2. As the vehicle travels ahead, the box area moves ahead the same distance. Any dust particle outside the box area is treated as dead. This method allows us to have a background texture which is not updated. So we only need to calculate and render the dust particles within this box area. This approach reduces the number of dust particles needed and also reduces the memory volume needed.

PARTICLES

For each calculated particle, we generate a number of particles that have movement similar to this particle, with some random behaviors. This way we only calculate one particle, but a system of particles will behave accordingly. It saves time for calculating all the particles and enhances the richness of the picture at the same time.

RESULTS

Fig. 6 and Fig. 7 are the simulation results under the same circumstances with different vehicle velocities (30 miles/hour and 60 miles/hour.) The number of particles and the density of the dust are greatly affected by the vehicle's velocities. Fig. 8 and Fig. 9 are the simulation results with different dust densities. There are many other parameter which affect the simulation. In most cases, the simulation looks better with more dust particles, but the simulation is much slower because all the particles must be calculated in the particle systems.

CONCLUSION AND FUTURE WORK

We have introduced our approach to simulating dust behaviors behind a traveling vehicle. We have two primary goals: one is realism of the simulation, the other is real-time computation. In order to achieve realism, we analyze the forces and factors and construct physically-based empirical models to generate particles and control the dust behaviors accordingly. In order to achieve real-time, we further simplify the numerical calculations by dividing the dust behaviors into three stages, and establishing simplified particle system models for each stages. We employ motion blur, particle blending, texture mapping, and other techniques in computer graphics to achieve better results. Our work is a useful addition to many applications in simulated virtual environments.

Our model is a physically-based empirical model. Using CFD to calculate the turbulent vortex behind a vehicle would yield an accurate physical model, which could be integrated into our work. The problem is that CFD models are too computationally complex. We have succeeded simulating fluid flows in real-time [1]. We hope to find a solution for dust behaviors also. Currently our vehicle is traveling in straight line. A logical extension will be to allow the vehicle to turn around. We are currently working on this.

We plan to further consider the interaction between the dust particles and the environment. For example, when the vehicle passes by, the grass on the two sides of the road will swing back and forth. We plan also simulate the dust accumulation on the grasses.

Figure 6: Vehicle Speed=30, Dust Density=40, BlurN=2

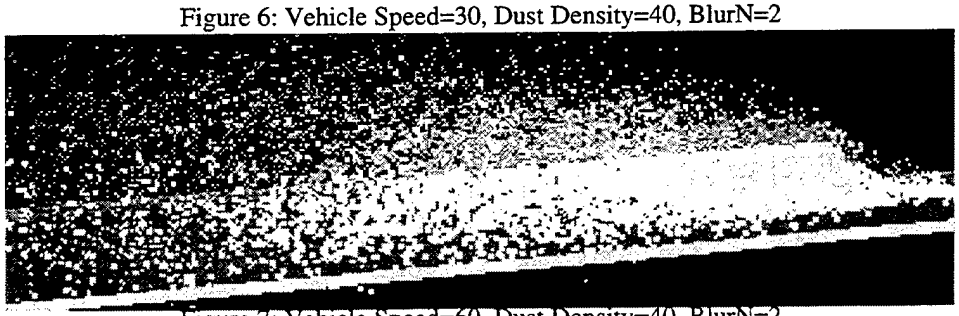


Figure 7: Vehicle Speed=60, Dust Density=40, BlurN=2

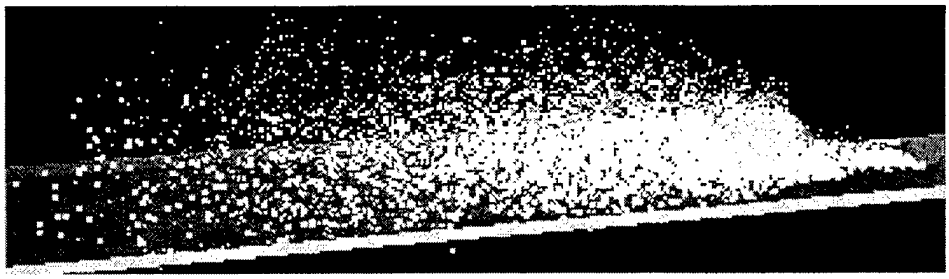


Figure 8: Vehicle Speed=30, Dust Density=10, BlurN=2

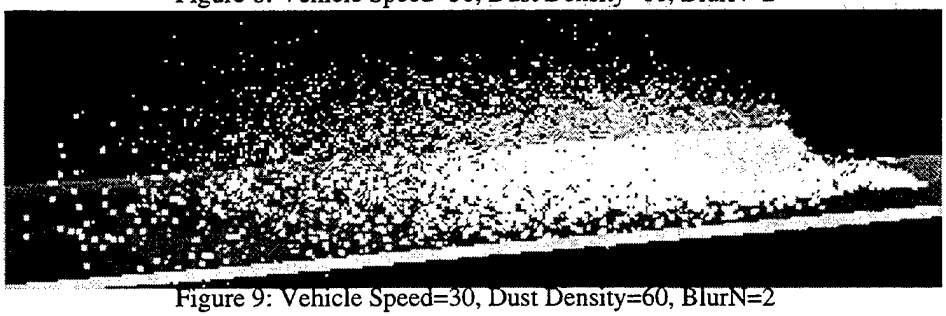


Figure 9: Vehicle Speed=30, Dust Density=60, BlurN=2

REFERENCES

1. J. X. Chen, and N. V. Lobo, "Toward Interactive-Rate Simulation of Fluids with Moving Obstacles Using Navier-Stokes Equations," *CVGIP: Graphical Models and Image Processing*, Vol.57, No.2, March, 1995, pp. 107-116.

2. C. Cowherd, M. Grelinger, and P. J. Englehart, "An Apparatus and Methodology for Predicting the Dustiness of Materials," *American Industrial Hygiene Association Journal,* v. 50 (Mar. '89) pp. 123-30.

3. S. Hsu and T. Wong, "Simulating dust accumulation," *IEEE Computer Graphics and Applications* v. 15 (Jan. '95) pp. 18-22.

4. 5. IST (Institute for Simulation and Training), *The DIS Vision: A Map to the Future of Distributed Simulation*, version 1, Univ. of Central Florida, Orlando, Florida, May 1994.

5. X. Li and J. M. Moshell, "Modeling Soil: Real-time Dynamic Models for Soil Slippage and Manipulation," *Computer Graphics* Proceedings, Aug. 1993, pp. 361-368.

6. T. Loke, D. Tan, H. Seah, "Rendering Fireworks Displays," *IEEE Computer Graphics and Applications*, May, 1992, pp. 33-43.

7. P. E. Oppenheimer, "Real Time Design and Animation of Fractal Plants and Trees," *Computer Graphics*, Vol. 20, No.4, 1986, pp. 55-64.

8. W. T. Reeves, "Particle Systems -- a Technique for Modeling a Class of Fuzzy Objects," *ACM Trans. Graphics*, Vol. 2, April, 1983, pp. 91-108.

9. P. Reffye, C. Edelin, J. Francon, M. Jaeger, C. Puech, "Plant Models Faithful to Botanical Structure and Development," *Computer Graphics*, Vol.22, No.4, 1988, pp. 151-158.

10. C. W. Reynolds, "Flocks, Herds, and Schools: a Distributed Behavioral Model," *Computer Graphics*, Vol. 21, No. 4, July 1987, pp. 25-35.

11. M. Roth and R. Guritz, "Visualization of Volcanic Ash Clouds, *IEEE Computer Graphics and Applications*, Vol. 15 (July '95) pp. 34-39.

12. K. Sims, "Particle Animation and Rending Using Data Parallel Computation," *Computer Graphics*, Vol24, No.4, Aug 1990, pp. 405-413.

13. J. Stam and E. Fiume, "Turbulent Wind Fields for Gaseous Phenomena," *Computer Graphics* proceedings, August 1993, pp. 369-376.

14. R. R. Williams and R. E. Davis, "Battlefield Dust Environment Symposium III," *U.S. Army Engineer Waterways Experiment Station*, Vicksburg, Miss., 1988. (Call No.: UG683 B39s 1988)

WAS EINSTEIN RIGHT?

By
Henry C. Alberts, PhD; Professor of Acquisition Management
Defense Systems Management College, Fort Belvoir, Va., 22060
Alberts_henry@dsmc.dsm.mil

ABSTRACT

This paper discusses attempts to apply standard analytical methodologies to three complex systems: (1) Field Experimentation mechanisms at the Army's Combat Development Experimentation Center at Fort Hunter Leggett [1962 – 1967]: (2) The "Small Independent Action Force" activities in Vietnam [1969 – 1971]; and, (3) The Defense Acquisition Process [1983 – 1994]. The discussion below describes each of the analyses performed and the results obtained. The data collected indicates that the key element to devising means of achieving better system performance and solving problems associated with such complex systems is flexibility to act as required, rather than construction of rigid rules and regulations. Taken as a body of information, the work suggests that in order to understand large, complex, man-machine systems, and to permit definition of generalized, long term solutions to perceived problems we must go beyond the use of classical analytical methodologies. It may be that finding long lasting workable solutions to problems may require new analytical mechanisms developed around the concept of bounded change rather than singular solution.[1]

BACKGROUND

Albert Einstein is cited as the author of the observation: "The thinking that got us to this point will not get us beyond it!"

For those of us who grew up in a simpler time, the greatest utility in classical statistics was that it helped us predict the likely outcomes of events about which we were unable to create deterministic models: that is, there was uncertainty about the problem, the variabilities which affected the situation, and, the relationships among them. Because statistics had often provided methodology to project outcomes in such situations, we extended its use beyond such clear cut, well bounded events as "coin toss" and "dice throws" to more complex, and less structured situations. Then, having invented multi-variate analysis and cluster analysis methodologies, we used them (along with our standard statistical techniques) to address more complicated equipment reliability issues, and extend our practice further into the realm of social system design.

We limited the scope of such inquiries to some extent by insisting on bounding the issues we did address. As Willis Willoughby (then head of reliability at NASA) said when

[1] Approved for public release; distribution is unlimited

he was presented with an analysis of the probability of sending a man to the moon and returning him safely, (of the order of .027), "That problem's outside the range of utility for standard statistical methodologies. – Let's get on with the job."

Today, "statistical methods" are used as analytical tools in examining extremely complex systems. The usual goal is to attempt to predict "outcomes" in situations where: (1) very large numbers of variables are involved; (2) we are not certain of the relationships among those variables; and, (3) human decision making provides the variables sleeted, their initial and subsequent values.

The following discussion describes some of the work that leads to the conclusions presented on page 6 *et seq.*

CASE HISTORIES

THE CDEC EXPERIENCE.

In 1962, I joined the Stanford Research Institute's CDEC Research Office as Chief of Instrumentation. My task was to devise, develop, install, and operate equipment capable of reporting the positions of everyone on the battlefield, and the events they experienced during experimentation. When the instrumentation was in place, we could analyze results and reach conclusions much more expeditiously. As we accumulated data from a number of field experiments, we noticed that there were elements of maneuver which were repeated in almost every one of them. Some examples were: (a) traversing a well known terrain area at different times of the day (or night); (b) simulating an attack on a high ground defended in one of several standard postures; or (c) sweeping an area to detect targets and their locations. We thought about creating a series of "set-piece" maneuvers; activities which could be performed repeatedly on selected terrains within the Hunter-Leggett Military Reservation, independent of a specific combat context. The idea was to construct experimental segments present in many of the combat field activities, replicate them sufficiently often to generate statistics about expected task performances, establish a data base of "standard field performance values for those combat tasks, and use the "statistical values" for those tasks instead of actually performing them in different experimental contexts.

We attempted to implement the idea. We planned a set of experiments specifically for the purpose of developing some performance statistics about how troops performed a set of specific tasks under controlled circumstances. We selected five "standard" combat tasks and structured a set of missions around them. We built scenarios which required each task to be performed. But in each scenario the order of task performance was varied. Since we did not have an unlimited number of troops, we formed minimum building block units and constructed larger units by aggregating them. Squads of five soldiers were used as the smallest building block. Platoons and Companies were built from squads. The larger units were asked to perform missions made up of a sequence of set piece combat tasks.

The scenarios were constructed to require task performance in different sequences. Finally, each scenario was replicated by the various mixes of troops in different organized units a sufficient number of times to provide a bounded value set for each combat task. Figure 1 shows the notional experimental matrix to be used in establishing the data base. In Figure 1, the segment column indicate the tasks to be performed, and the sequence columns indicate the order of task performance.

Figure 1
THE CDEC "SET PIECE" NOTIONAL MATRIX

	PERFORM SEQUENCE #1	PERFORM SEQUENCE #2	PERFORM SEQUENCE #3	PERFORM SEQUENCE #4
SEGMENT #1	1	2	3	4
SEGMENT #2	2	3	4	5
SEGMENT #3	3	4	5	1
SEGMENT #4	5	1	1	3
SEGMENT #5	4	5	2	2
↓	:	:	:	:
SEGMENT #n	A	B	C	D

But when experiments containing "set-piece" elements were performed within a combat context, the data differed considerably from what had been anticipated as a "standard data set". It became very clear that (a) mission context had great effect on unit performance; and (b) terrain, mission context, and tactical assumptions of friendly and enemy force distributions formed sets of complex relationships which appeared to provide unique situational results.

A contributing cause of result non-reproducibility was thought to arise from the nature of human beings. They learn from their experience. As Kant philosophized [1], and Damasio later observed experimentally and noted in his book "Descartes' Error" [2], we tend to see things within the context of our own experiences and the way we see them has an emotional component which determines how we record the experience in our brain. When we observe a segment of mission performance, what we observe depends upon all of the life experience precedent to that segment of activity. Thus, a major component of the effect we observe in any mission performance depends on our individual prior history - which includes the order of previous task performance, and our individual sense of mission continuity generated by those prior experiences as well as the emotional component of the incidents recorded.

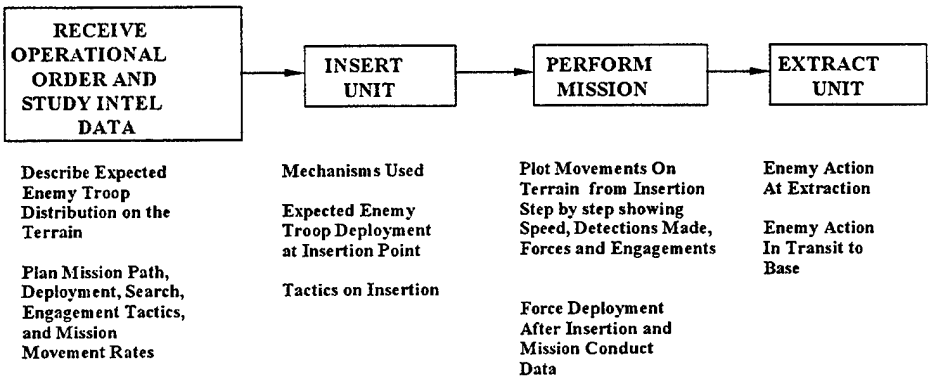
THE SMALL INDEPENDENT ACTION FORCES (SIAF) EXPERIENCE

In 1968, the Defense Advanced Research Projects Agency (DARPA) authorized construction of a data base about activities of Small Independent Action Forces (SIAF)

operations in Vietnam. Analysis of SIAF operations had shown that small patrols of from 5 to 9 troops operating for time periods of between 2 and 9 days exterior to larger organizations (companies, or battalions) were more successful in performing their missions than were similar sized units operating within, and under the control of larger force elements. Compared with traditional larger force elements, these units had fewer casualties, expended less ammunition, and reported higher numbers of enemy detections while remaining undetected themselves. There were four types of small units involved: (1) SEALS (Navy units), (2) Long Range Reconnaissance Patrol (Marine units; (3) Ranger units (Army); and (4) Special Forces units. The idea was to determine how small units made operational decisions during patrols with emphasis on how perceptions of enemy forces influenced those decisions. By doing so in both combat and non-combat situations, we hoped to provide a data base permitting construction and test of a combat model which would validate hypotheses based on lessons learned in the field. Over the period of 16 months, our team reviewed 1,500 SIAF "after action reports" and collected data from an equal number of patrols. All of the services were about equally represented.

Our first finding was that SIAF units had developed the mission time line shown in Figure 2.

Figure 2
SIAF UNIT MISSION TIME LINE



In Figure 2, boxes surround the each element of the SIAF mission. The data elements collected about each mission element are listed below each mission element box.

When we collected data from returning patrols, we met the patrols as soon as they had finished their "mission completion chores". Using the model in Figure 2, we took minute by minute data about each patrol from the time they were alerted to expect an operational order, to their withdrawal from combat terrain. We were interested in how patrol routes were selected, speed of terrain traverse, reconnaissance methodology, target detections, tactical choice rationales and almost all of the minutiae embedded within the combat mission sequence. We also collected a full set of demographic characteristics about the individual squad members and their unique histories. At the end of our 12 month

in country data collection activity, we felt we could construct a data base which we represented adequately the sum of experiences of small forces in Vietnam combat.

To test the reproducibility of the SIAF experience in areas other than Vietnam, we established a field test site in the Hawaii National Forest on terrain similar to that reported by SIAF units in Vietnam. On the experimental terrain, we measured parameters of SIAF type performance achieved by Vietnam-veteran Army and Marine squads operating in the SIAF mode. 23 patrols operated within a set of simulated combat situations. We measured the patrol performance with special instrumentation borrowed from CDEC. We collected the same data about participants as we had collected in Vietnam. When we compared the set of test data from Hawaii with the Vietnam Interview data, we noted that velocity of movement over terrain for various kinds of terrain, and the ratio of target detection to numbers of target available for detection were heavily dependent upon the specific mission experience as it unfolded, and secondarily on terrain parameters.

We then built a computer assisted game to use at the Army Special Forces School. Using Vietnam combat veterans as test subjects, we would simulate a SIAF patrol in a closed environment. Simulations began upon delivery of a mission operational order. Data was taken about how mission details were planned and executed. Pictures of the actual mission terrain were used to help patrol leaders make key decisions about movement and force tactics on that terrain. At each step of the way, the computer compared decisions and actions of the test troops with the data derived from Vietnam and from the test facility in Hawaii.

When we compared the three sets of data, we reached an interesting set of conclusions. In general, performance varied with insertion method, order of mission task performance, perception of distributed enemy forces, and perceived operating environment (terrain, weather, stimuli, etc.). Details of unit performance were not duplicated even in similar environments, and when proceeding under similar operational orders. The specific findings were

(1) for each of the three data bases, the decision rules which led to the observed results were different;
(2) the consolidated data base inclusive of all three of the independently derived data sets could be bounded; but,
(3) while a bound could be defined for the measured parameters (rate of movement over terrain, target detection ratio, range of detection, etc.), that bound was strongly dependent on the patrol's perception of: (a) enemy troop distribution; (b) numbers of enemy detections made from insertion to the time at which measurements were made; and (c) the terrain characteristics;
(4) the boundaries which resulted from each data set result (i.e., the probability density functions) were different because the number of variabilities embedded in the observations were so large and depended so heavily on individual troop perceptions; and,

(5) because of the we had only a moving boundary and could not suitably integrate the data into a singular data base, we were unable to use the results to predict narrow bounds for future performance by similarly configured troops comprised of equal numbers of individuals with similar demographic, training, and combat histories.

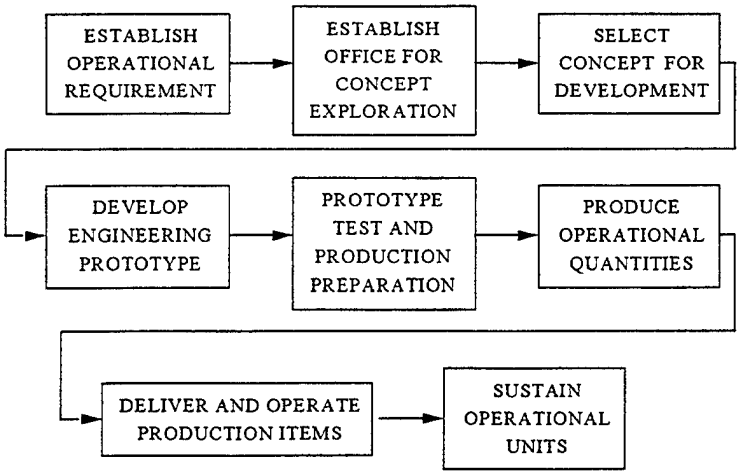
We also made some conjectures:

(1) task performance may depend on decision chains: (short and long term learning effects) plus perceptions of what is, and has been, going on during the mission, test/experiment, or game, and
(2) performance of the same mission sequences may not be reproducible in the same/similar environments.

THE DEFENSE ACQUISITION PROCESS

In 1983, when I became Professor of Engineering Management at the Defense Systems Management College, I began to study the process by which the Services purchased major weapon systems. The process at that time was described by the old Department of Defense Directive 5000.1. It proceeded from the establishment of an operational requirement through the delivery and operation of end weapon items to the fighting forces. Figure 3 shows the major steps in the process.

Figure 3
STEPS IN THE DEFENSE ACQUISITION PROCESS



Our initial plan had been to: (1) collect detailed anecdotal data on effect of perturbing events on acquisition process performance; (2) form a cause/effect data base of relationships and their effects on cost and time-line projections; and (3) Compare data base results with programs external to the data base to establish data utility. To implement

the plan, we began to gather statistics on elapsed time between: (1) definition of an operational requirement and formation of a group to explore competing system concepts; (2) establishment of the concept exploration group and selection of the concept to be developed; (3) selection of the concept to be developed and completion of the engineering prototype; and (4) narrative descriptions of events which affected those elapsed time intervals.

We observed a broad range of elapsed time intervals for each of the acquisition process steps. When we explored the "reasons" why there was such performance diversity, we noted that many could be traced to decisions during process execution taken by those outside the direct acquisition activity chain.

One major contributing factor involved how resource streams were allocated to the program. Specifically, gaining approval to proceed with concept exploration requires provision of pro-forma "program budget" and "time to complete" estimates for the entire development program (which might be expected to take up to 15 years). These estimates almost always assumed that: (1) resources would be available as required; (2) the concept/engineering/production/sustainment activities as stated were those which would ultimately be proven correct; and, (3) there were no unforeseen impediments either to the order in which all of the activities described were to be performed, or the time line projected for their performance.

However, it was almost always the case that: (1) actual resource levels failed to meet planning (both in amount available and the dates of availability); and (2) the lack of timely resource availability or some previously unknown technical problems arose which prevented the forecasts from being implemented as planned in the initial assumptions.

It also became clear that the point in time at which program resources became constrained, or technical assumptions proved to be in error were major factors in determining what happened along the time line from initial program authorization to ultimate operational deliveries. The effects observed on the outcomes of each succeeding process step were different depending mainly on those two factors. While it was possible to compute a boundary within which all observed times for step accomplishment fell, we did not know how to use that information to predict what would happen if resources were constrained by some arbitrary amount at some random point along the time-line of program execution. There were a great number of additional variables which depended upon the: (1) rate of change of the end requirement; (2) kind of program being pursued; and, (3) rate of technology advance during the development period. In short, there were so many variables involved that we simply could not collect data in sufficient amounts to permit us to do good statistical analysis. We decided to try another methodology.

With approval, funding, and active assistance from the Undersecretary of Defense, Acquisition, we held a series of three day workshops to explore in detail the problems which affected the acquisition process for smart munitions and other kinds of weapon

systems. Each workshop had 10 participants. All participants were Program Managers or Deputy Program Managers from government (military and civilian) and contractors. Their senior staff members were available to provide additional details of pertinent information. During each workshop, participants would describe details of the perturbing events they experienced, the responses taken to those events, and effect of the events on program performance experienced.

As we continued to explore the process. we found great diversity in the types of problems experienced during program performance; and we derived relationships among them. At the conclusion of our work, we created a data base from all of the workshop data. The data base enabled us to produce a relational model that showed us how problems experienced at some point in the process affected the outcomes at succeeding points along the time line of process performance. These relational models provided us with a good understanding of how an event might influence process performance in the context of the particular situation. But we also discovered that the same event (i.e., a budget cut of 10%), occurring at a different time in the process (i.e., concept exploration vs. engineering development); in different programs (i.e., a ground to air missile or an air to ground missile) even among those using the same technology, could produce very different results in terms of completion dates and costs for succeeding process steps.

SOME TENTATIVE CONCLUSIONS

We established some elements of similarity in each of the cases discussed above:

(1) All of the processes (i.e. experimentation at CDEC, small force combat in Vietnam, and exercise of the Defense Acquisition Process) were complex., and were influenced by many variables which, in turn were of complex structure.

(2) Process steps generally were describable as instances of repetitive task performance; the processes were made up of a set of fixed functional sequences; but the functional sequences were performed within the context of different complex situations: situations which themselves were composed of many inter-related factors capable of influencing the process outcome.

(3) The temporal relationship between events which affected the process outcome and the point along the process time-line at which those events occurred was variable and often depended upon factors not necessarily embedded within the process.

We found that in the situations discussed above, taking exactly the same action (e.g., absorbing a budget reduction of 10%) in different contexts (e.g., a "smart munitions" program or a new aircraft development program), and at different points along the process time-lines (e.g., in concept exploration or engineering development and test) produce situationally dependent singular effects.

The findings across all three cases led us to consider the question of whether applying classical mathematical methods to these kinds of issues requires making some fundamental assumptions unlikely to be present in the real world. For example, is it realistic to assume that: (1) there is "event reproducibility" across time and space; (2) the population of outcomes is bounded or stable; or, (3) there is "central tendency" in repeated sampling. Where there is absence of repeatability; when the dimensions of outcomes are unbounded; or when the numbers of samples required to generate acceptable levels of confidence are very high, we might re-consider the use of classical statistical approaches. It might be necessary to develop new mathematical concepts better suited for addressing the kinds of issues involving large, complex systems which incorporate a multitude of human-machine interfaces.

When such conditions are present it might be more appropriate to use perturbation and/or chaos concepts than to try and satisfy requirements for application of standard statistical methodologies so useful in situations which can be replicated reasonably well. In short, does the richness of situational and process complexity make each result observed non-repeatable, even though an effect observed can be assigned to what appears to be the same cause (a 10% funding reduction). We believe that this kind of problem construct can make it necessary to find limits of cause and effect rather than equations which link discrete cause to particular effect.

Perhaps a good illustration can be found in Gleick's discussion of Foucault's Pendulum [3] which points out that while the exact path of the pendulum is not predictable, the boundaries within which that path will lie (i.e., trajectory bounds) are; and the focus around which all possible paths will reverse for a bounded set of initial and subsequent conditions (the "attractor") can be predicted While Gleick's illustration is relatively simple, there is no difficulty in finding additional complex situations in which one would like to predict an outcome. However, such situations appear to be characterized by non-repeatable "cause and effect" linkages within contexts of large numbers of variables which can effect the outcomes.

There may be yet another difficulty in applying some standard statistical methodologies to complex processes: interactions among complex processes and complex environments may preclude the assumption of a "central tendency" in observations made over time. There may even be difficulties in application of chaos concepts in such circumstances. It may be that, over time, the location of the attractor in the complex space of its existence, changes in response to changed values of parameters which determine its location. Such occurrence would preclude specifying fixed attractor loci. Thus, a "space within which" points will be found (with no preference about where in that space particular effects of particular causes will lie) might be all that can be postulated for each iteration of a particular complex process interacting with its environment. It may also be that as complex processes play out, the boundaries of attractor space change to accommodate changes in context (relationships among system components) of the system under study and the environment with which it interacts and is affected.

One possible illustration of this kind of situation may be the efforts of Motorola to achieve, and then maintain a reduction in its manufactured component rejection rate by tightly controlling the manufacturing processes such that "a six-sigma boundary" is experienced for all of its various manufacturing operations. In describing what was done and how it was accomplished, Motorola representatives said that while the effort had been very successful in providing tightly controlled manufacturing processes, they had also observed that the point about which the tight control was held (the mean value) would shift over time. They said that the shifting mean could not be assigned simply to changes in process tooling or to change in materials input to the process. Those parameters were capable of tight control. To the question, "What is causing the shift in mean value when the boundaries of repeatability are maintained so well?", the response was: "We don't know and haven't been able to find out!" One can hypothesize that in this case; (1) the attractor is the mean value of items produced by the process; (2) the boundary of possible values is six sigma; and, (3) the shifting mean is the system's response to change to the complex environment within which the process operates.

There is a very real problem which arises if one cannot depend upon the analytical results drawn from observations of complex systems over time. It is linked to the way we have tended to "solve" problems in the past. Increasingly, we have attempted to make sweeping sets of constraints on allowable behavior when certain situations arise, (e.g., a law, or regulation). ***Such prohibitions create an inflexible set of boundaries which limit response even when the spectrum of allowable responses is inappropriate to the situation***.

The issue is not simply one of finding a set of appropriate responses in a static or fixed bounded situation; it is rather a problem of attempting to take appropriate action under conditions of change when the time constants of change for the system and its environment may be considerably different from the time constants of change for the system of constraints. The Defense Acquisition process is a good case in point. Title 10 of the United States Code provides the governance for how the acquisition process is carried out by members of the Defense Acquisition Corps. The Congress is responsible to make change to the U.S. Code through the Legislative process. In the absence of some very compelling reason, up to a year can elapse before a change can be made even when there is a feeling of urgency about the need to make change. The legislative mechanisms are designed for deliberative process, rather than for emergency actions. Emergency action rules are provided for in the operating regulations which result from legislation: but regulations are also slow to change even under extreme conditions of perceived need. It has taken more than 3 years to make any major changes in the acquisition process!

In many instances, situations have arisen over the years which indicated the need to take actions forbidden under then existing rules and regulations. At those times, Acquisition Corps members had some choices: (1) doing something not really appropriate to the situation, but within the boundaries of permitted actions; or (2) doing something

"outside the rules" and taking action they believe appropriate in that specific situation. The dilemma between doing what is permitted and what is deemed "appropriate under extant regulation" can create considerable strain on the acquisition work force. Often, stepping outside the boundaries of permitted action may resolve the immediate issue, but it lays the individual open for post facto analyses in which breaking rules becomes more important than a successful outcome. If the outcome has been less than successful, the result of such inquiry is likely to be very bad for the individuals involved.

The presumption that a body of statistical data can be developed which can provide the basis for structuring sets of "best appropriate actions" under all circumstances " has not been readily demonstrated in these three instances.

A summary of the conjectures resulting from this work can be stated as follows:

- The processes studied in this work are sufficiently complex to make data reproducibility doubtful.
- Outcomes appear to be singular and to depend heavily on decision chains influenced, but not predetermined, by perturbing events.
- Like magnitude perturbations in similar circumstances produce different outcomes depending upon situational context.
- In such processes, with their large numbers of independent variables, and the uncertainties of variable dependencies, cause-effect linkages appear to be confounded.
- Gaining sufficient data to develop confidence in predictions made from the data base may require a long time period.
- If long time periods are required for data collection and analysis, the environment within which the process operates may undergo considerable change.
- Half-life of collected data and the results obtained from it may be too short to test data base utility or prediction validity.

A SUGGESTED FUTURE COURSE

Suppose one were required to develop an approach to understanding complex systems in the world of non-repeatable cause-effect observations where a central tendency has been replaced by a temporary, changing central value which depends upon a continuing stream of changing relationships in the system/environment/context paradigm. How could one proceed? In a world which is rapidly becoming more and more interconnected, it seems reasonable to predict that future problems will be characterized by increasing numbers of interacting, rapidly changing variables which impact on the way we live, conduct our business, and make decisions for our future. Under those conditions, there will be greater need for tools which permit treatment of complexity in a timely, yet thorough manner.

The statistics of the future will require consideration of a set of new concepts

founded on characteristics of non-repeatable events which take place within changing environments. It may also be necessary to provide entirely new mechanisms to integrate and display the data which are necessarily involved in those situations. Even today, human beings are reaching the limits of cognitive capability when dealing with complex issues described by large amounts of information presented in traditional ways (e.g., matrices, or linear models). As the need grows to understand and integrate larger amounts of information, the limits of cognitive capacity will be stretched perilously.

The task for those of us who strive to develop and apply analytical tools to difficult issues, is to consider expanding well beyond the envelope of traditional methodologies and finding ways to accommodate the needs of our future. In such a quest, we can again look to Einstein and ask whether he has anticipated the problem. He is known to have said often that:

"Imagination is more important than knowledge!"

ACKNOWLEDGEMENTS

I am indebted to the over 1400 individuals who contributed their knowledge and understanding to the investigations reported here. Without their contributions, there would have been little to write about.

REFERENCES

[1] Kant, Immanuel; "A Critique on Pure Reason"; 1990 ; Promethius Books, Buffalo, N.Y. 14228-2197; ISBN 0-87975-596-2

[2] Damasio, Antonio: "Descartes Error" 1994; Avon Books, New York, N.Y., 10019; ISBN 0-380-72647-5

[3] Gleick, James; "Chaos"; 1987; Viking Press, New York, N.Y., 10010; ISBN 0-670-81178-5

Henry C. Alberts
Fort Belvoir, Virginia, New York, N.Y.,
and Barrington, R.I.
22 December 1997

Predictive Statistical Process Control for the Military

David H. Olwell *
Department of Mathematical Sciences, U. S. Military Academy
West Point, NY 10996-1786

November 10, 1998

Abstract

Olwell(1997b) developed methods for the control of processes which have low frequency events, usually failures. Here we illustrate the methods by considering three types of isolated incidents: Driving Under the Influence (DUI) offenses from an armored division in Germany, Class A aviation accidents in the Department of Defense, and rapes in New York City. We then propose that these methods be adopted for key command interest items in the military. The implications of this proposal are briefly discussed.

Keywords: Predictive distributions, short run, Poisson, military, accidents, rape.

1 Introduction and Background

Statistical Process Control (SPC) is a body of graphical techniques which separates usual variation from unusual variation in a process, based on a model. We will consider two types of charts in this paper, which identify isolated departures from the model and persistent departures from the model. These methods are self-calibrating: they learn from and adjust to the historical record, and can be implemented without the large training sets of data that characterize classical SPC.

We propose to apply these methods to the management of sexual harassment data. No such data is available in useful form from military sources. To illustrate the methods, we turn to three data sets believed to be similar in nature to data

*This research was supported by the Army Research Laboratory, the USMA Mathematical Sciences Center of Excellence (MSCE) and the Dean's Fund for Faculty Research, USMA. Lieutenant Colonel Olwell is the Director of the MSCE, and an associate professor at the Military Academy. Approved for public release; distribution unlimited. "The views expressed herein are those of the author and do not purport to reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense."

on sexual misconduct. They are rapes in a precinct in New York City; driving while under the influence (DUI) offenses in an Army division in Europe; and aircraft Class A aviation mishaps. These examples illustrate the methods and their generality, and provide scenarios for further illustrating how the methods are useful for turning data into meaningful information for the commander.

The technical underpinning of these methods was presented in an earlier article (Olwell, 1997b). Here we focus on the applications and analysis.

The military has a need for methods like this which can be applied to a range of problems, including but not limited to regular misconduct, sexual misconduct, accident rates, equipment failures, and suicides. In each case, leaders need to know if the variation in the processes can be classified as usual or unusual. Traditional SPC methods frequently do not apply, because the extensive historical data needed to calibrate them is not available; or because policy changes make previous historical data less relevant to current processes. When investigation of sexual misconduct in the Army began, for example, there was no useful historical data available to calibrate traditional charts. Self—starting charts and methods such as these were needed.

Similar situations occur in industry and other levels of government.

Leaders find these charts easy to understand and interpret, as our experience with both uniformed and civilian military leaders and with the leaders of the New York City Police Department has shown.

As a consequence of the Aberdeen scandals of Fall, 1996, the Army has been revisiting its methods for reporting and monitoring misconduct especially sexual misconduct. There is no central reporting mechanism as of this writing. Should data be collected, the methods of this paper were specifically designed for monitoring those reports and detecting both unusual reporting periods and persistent changes in the reporting process.

No incident of misconduct, sexual or otherwise, is acceptable. It represents a failure, and must be addressed. However, the Army has procedures in place to deal with its usual levels of misconduct. Senior leaders who monitor these procedures need tools to distinguish usual levels from unusual levels, and to detect isolated or persistent changes from historical levels. Leaders hold extraordinary measures "in reserve," and must commit them wisely. It cheapens the effect of extraordinary measures to apply them when they are unwarranted. Worse, the measures themselves can cause shocks to the underlying system and destabilize it. On the other hand, failing to react to extraordinary events may allow a bad situation to become much worse. This dilemma is analogous to the decision to use an antibiotic: its frequent indiscriminate use markedly decreases its effectiveness, yet its omission may result in the death of a truly sick patient.

The methods of this paper provide methods for determining when to implement extraordinary measures. We illustrate this point with our data sets, especially with the Class A aviation mishap data from 1997.

In other words, we are arguing to apply the statistical thinking underlying the Deming philosophy to the management of data used to describe key command interest items in the military. Our contribution is to have developed methods which work for the application of Deming's ideas to short–run Poisson

data.

2 Example — Class A Aviation accidents

We turn to four examples to illustrate the methods. We note that there is no useful data on incidents of sexual harassment in the Army extant, at least for the purposes of these methods. We turn to similar examples. We examine aviation accidents, rapes in New York City, and lastly misconduct (albeit DUI) in a division in Germany.

2.1 Example — Army Class A aviation mishaps

A Class A aviation mishap in the military is one that involves a fatality or damage in excess of one million dollars. For Fiscal Year (FY) 1997, the number of Class A aviation accidents per week is given in Table 1. The weeks run from Sunday to Saturday, so 53 weeks are reported. We are interested in detecting unusual isolated weeks and persistent shifts from the usual accident rate.

We note that this data would better analyzed by weighting for the number of flying hours each week, but that data was not available to the author.

We will first apply the methods of this paper to the Army aviation data set, first with a weak prior and then with a strong prior. For our weak prior, we choose $\alpha = 1$ and $\beta = 4$. This corresponds to an expected rate of 0.25 incidents per week, with a great deal of uncertainty as to the true value.

The charts for an isolated departure, persistent departure, and the posterior distribution at the end of the year are shown at Figures 1, 2, and 3. They behave much more reasonably. Interpretation of these charts shows that the first four weeks with accidents are signaled as unusual. After the first twenty weeks, enough evidence has accumulated about the underlying accident rate to see that one accident in a week is not an isolated model departure.

We have set the persistent chart with a two–sided average run length of 250. Against that criteria, there is not evidence of an upward shift, although a less sensitive threshold would signal an upward increase in the mean number of accidents.

We compare these charts with those obtained using a strong prior, say $\alpha = 10$, and $\beta = 40$, shown at Figures 4, 5, and 6. This prior corresponds to almost a year's worth of data. We see that the isolated chart (which now looks exactly like the original data plot) no longer displays a learning period, and that there is no evidence of a persistent upward shift.

2.2 Example — DoD Class A aviation mishaps

This example considers the Class A aviation mishaps for the entire Department of Defense for Fiscal Year 1997, with the data by week. The data was provided by the Assistant Deputy Under Secretary of Defense for Safety and Occupational Health, and appears in Table 2.

Week	Number	Week	Number
1	0	28	1
2	0	29	1
3	0	30	0
4	0	31	0
5	0	32	0
6	0	33	0
7	0	34	1
8	0	35	0
9	10	36	0
10	1	37	1
11	0	38	1
12	0	39	1
13	0	40	0
14	0	41	1
15	0	42	0
16	1	43	0
17	0	44	0
18	1	45	0
19	0	46	0
20	0	47	0
21	0	48	0
22	2	49	0
23	0	50	0
24	0	51	0
25	0	52	0
26	0	53	1
27	0		

Table 1: Data on the number of Class A aviation accidents by week for the United States Army, Fiscal Year 1997. Source: Assistant Deputy Undersecretary of Defense for Safety and Occupational Health, US Department of Defense.

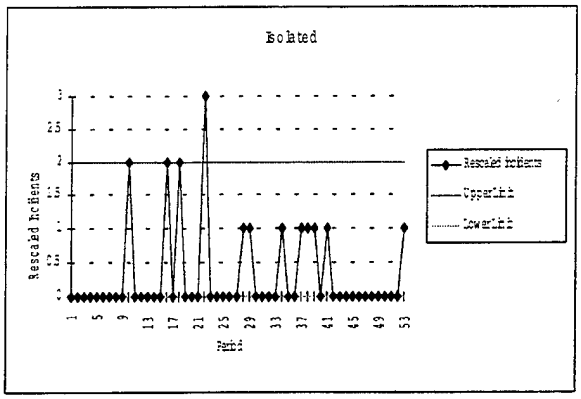


Figure 1: Predictive control chart for isolated departures with a weak prior applied to the Army aviation data.

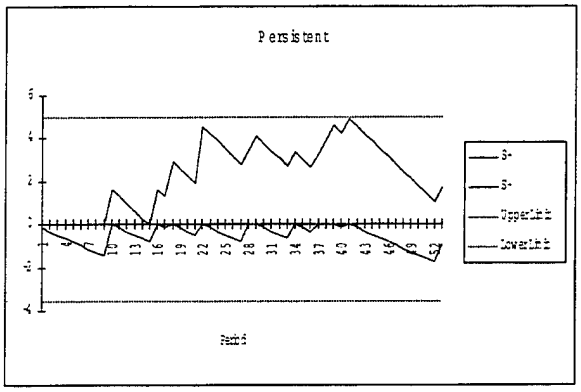


Figure 2: Predictive cumulative sum control chart for persistent departures with a weak prior applied to the Army aviation data.

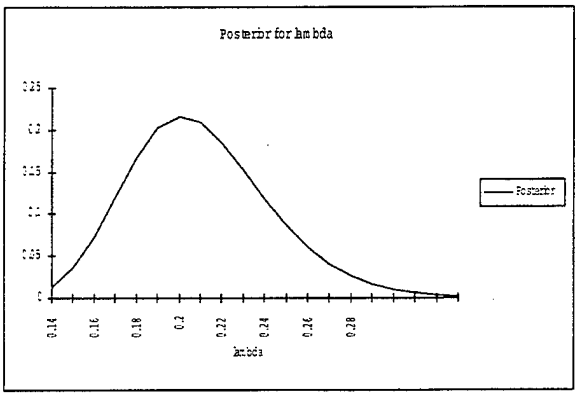


Figure 3: Posterior distribution for λ after week 53 for the Army aviation data, with a weak prior distribution.

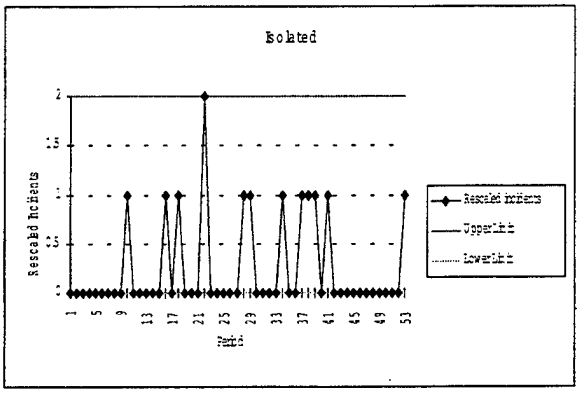


Figure 4: Predictive control chart for isolated departures with a strong prior applied to the Army aviation data.

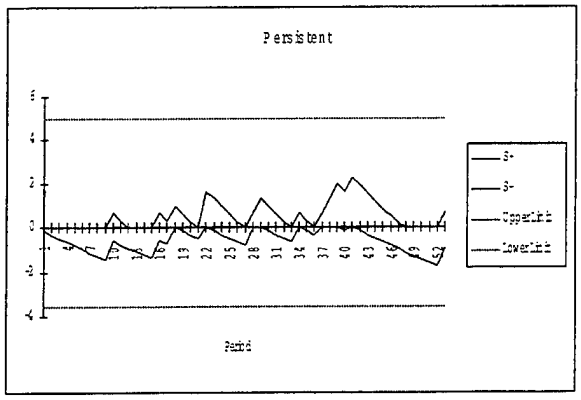


Figure 5: Predictive cumulative sum control chart for persistent departures with a strong prior applied to the Army aviation data.

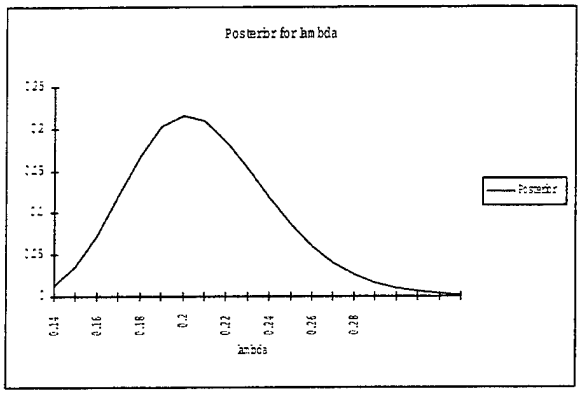


Figure 6: Posterior distribution for λ after week 53 for the Army aviation data, with a strong prior distribution.

Week	N	Week	N	Week	N
1	0	19	2	37	1
2	2	20	2	38	1
3	2	21	2	39	3
4	1	22	3	40	0
5	1	23	0	41	2
6	0	24	1	42	0
7	0	25	1	43	0
8	2	26	1	44	1
9	1	27	3	45	1
10	3	28	1	46	2
11	1	29	5	47	4
12	0	30	1	48	0
13	0	31	1	49	1
14	0	32	1	50	1
15	3	33	1	51	5
16	1	34	2	52	1
17	0	35	1	53	2
18	2	36	1		

Table 2: Class A aviation incidents by week for FY97, with overlap since the weeks run Sunday to Saturday. Source: DOD.

We note some obvious shortcomings in the data before we proceed to analyze it. First, it seems reasonable that the expected number of accidents would be proportional to the number of operating hours. Unfortunately, that data on operational hours was not provided. As a result, the analysis is not able to account for decreases due to reduced flying hours (such as usually occurs during the Christmas–New Year's holiday period.) The data does seem to be reasonably well fit by the Poisson distribution, however, based on the Poisson dispersion test.

We proceed to analyze the data, obtaining the Figure 7 and 8.

Analysis of these graphs indicates two weeks (29 and 51) with unusually large numbers of accidents. The second week, in September of 1997, coincided with increased press attention on military aviation mishaps. There was no evidence of a persistent increase in the underlying accident rate, however.

This raises the issue of what an appropriate response to an isolated bad week should be. Is a DOD–wide stand down appropriate for an isolated week? Or should that extraordinary corrective measure be husbanded and used only when there is evidence of a persistent problem? Why was there not a stand down in response to week 29's accidents?

The answers to these questions are informed by the analysis that the methods of this paper provide.

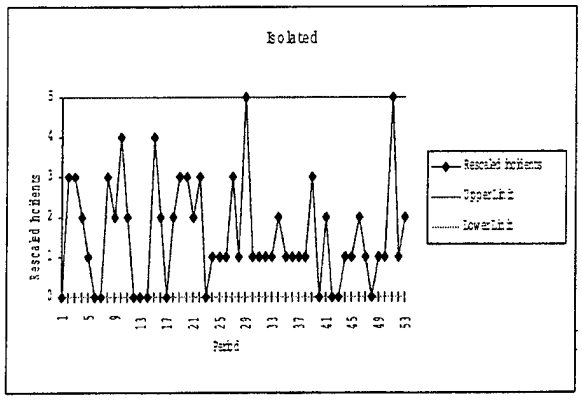


Figure 7: The chart for isolated departures for the Class A aviation data. Notice that weeks 29 and 51 have unusually high numbers of incidents.

3 Example — Rapes in New York City in 1996

3.1 Setting

Our second example is provided by the New York City Police Department. Crime statistics for each precinct are collected weekly. We are interested in the number of reported rapes each week for a given precinct. We assume that the distribution of reported rapes follows a Poisson distribution. (A test of this assumption was made on the entire data set, using the "Poisson dispersion test" referenced in Perry and Mead [1979], with p–value 0.225.) We set $\alpha = .02$, roughly corresponding to one false alarm per year while in control. We use a vague prior initially.

The population of the precinct is assumed constant, as is the length of the reporting period. For ease, we have set $n_i t_i = 1$, so λ is the mean number of arrests per week in the precinct.

The number of reported rapes is provided in Table 3.

3.2 Initial charts

The predictive control chart is included at Figure 9, and indicates that week 24, with six rapes, had an unusually high number of rapes. This would be cause for investigation by the precinct commander and his staff.

The posterior distribution of λ, after the 52 observations, is at Figure 11. Notice there is still a fair amount of uncertainty about the true mean rate of rapes. The point estimate for $\hat{\lambda} = 2.28$. The posterior distribution for λ, after only 22 weeks, is given at Figure 12. After 22 weeks, the point estimate for $\hat{\lambda} = 1.9545 = G_{22}/H_{22}$.

Week	N	Week	N
1	1	27	2
2	1	28	4
3	3	29	3
4	2	30	2
5	1	31	4
6	2	32	3
7	5	33	4
8	0	34	2
9	1	35	2
10	1	36	4
11	5	37	1
12	4	38	2
13	1	39	1
14	2	40	1
15	3	41	4
16	2	42	4
17	2	43	0
18	2	44	2
19	2	45	0
20	1	46	3
21	1	47	4
22	1	48	2
23	4	49	3
24	6	50	1
25	3	51	2
26	2	52	1

Table 3: Number of reported rapes by week in the 75th Precinct, New York City Police Department, for 1996. Data provided by Detective Josef Falletta, NYCPD.

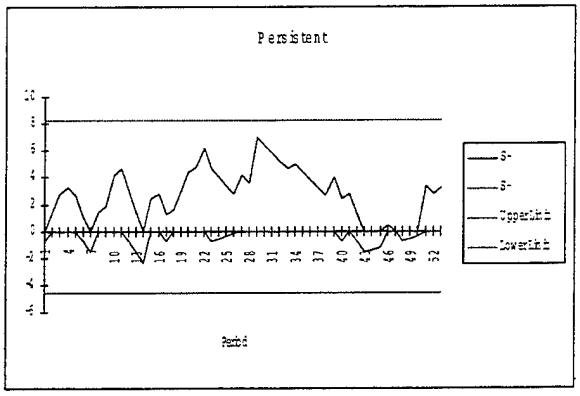


Figure 8: The chart for persistent departures for the Class A aviation data. There is no evidence of a persistent shift in the average number of accidents per week.

3.3 Restarting the charts

We illustrate in this section the ease of restarting the charts.

Since we have warmer weather, an indication of a step change after week 22, and an unusually large observation at week 24, we restart both charts from week 23. We have some history about the process from the first 22 weeks, so we will use an informed prior.

We believe that the new λ may be $m = 2.5$, plus or minus $s = .5$. We obtained this estimate from considering the increased average of weeks 23–28, but discounting week 24, which was possibly an outlier. We are unsure about the precision so we use a large value of s. Matching moments, we obtain $\alpha = 25$ and $\beta = 10$. We use these new values for our prior distribution when we restart the charts.

We could have validly held other beliefs about the parameters.

For the rest of the year, we obtain the charts at Figures 14, 13, and 16. At the end of the year, we see that the new point estimate for λ is given as $\hat{\lambda} = 2.469$, an increase of 26% over the rate for the first 22 weeks.

In Figure 15, we see that there is no further step change in the reporting patterns.

From Figure 13, it appears we underestimated the new value of λ, but the chart adjusts quickly to a value closer to 3 than to 2.5.

3.4 Analysis and comments

The rate of reported rapes appears to have increased in week 23, and then held steady for the rest of the year. This suggests that there may be a seasonal component in the rape rates, related to the summer months. However, analysis of the city-wide data shows no corresponding city-wide increase, which eliminates

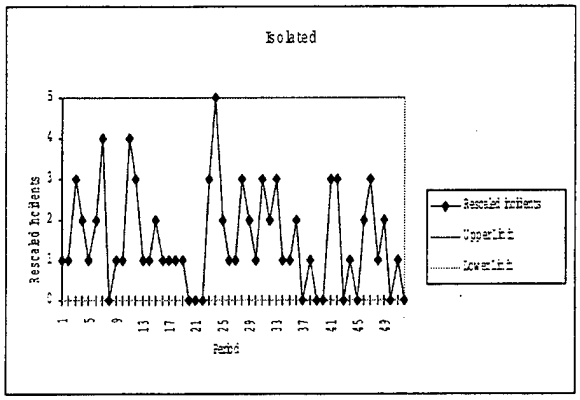


Figure 9: Predictive Control chart for isolated departures from the expected number of rapes per week in the 75th Precinct, NYCPD. There is only one signal, at week 24.

the seasonal argument. Additionally, the city as a whole experienced a persistent and marked decrease in the reports of rape starting at week 40, and this precinct did not experience such a drop. We discuss this in the next section.

The signal in week 24 may be due to the increase in λ, not an isolated special cause.

These methods only signal that there has been a change. They do not explain why — but they do give an indication of when the shift likely occurred, which is very useful for diagnostic work. The commander wants to know why shifts occurs. Modeling λ as a function of other covariates – such as weather, overtime, number of officers on duty, unemployment – would be a useful exercise in Poisson regression. This would provide additional useful information to the precinct commander.

3.5 City–wide charts

Although not developed in this paper, Olwell (1997a) has developed similar predictive control charts for the normal distribution. The self–starting charts of Hawkins (1987) (Hawkins and Olwell, 1998) can also be used, and are equivalent to the predictive chart when the predictive chart has a vague prior distribution. We assume a vague prior, and consider the city-wide data. The sum of all the rapes in all of New York's precincts is well–modeled by the normal distribution, both for empirical and theoretical reasons. The charts for the mean for the city as a whole is shown in Figures 17 to support the earlier assertions that the phenomena affecting the 75th precinct are local and not city wide. From the chart we see that there was no city wide increase at week 23, when the 75th precinct was experiencing an increase, and there was a city wide decrease beginning at week 43 which was not reflected in the Precinct.

Comparisons such as this would be useful for the Army. The analog is

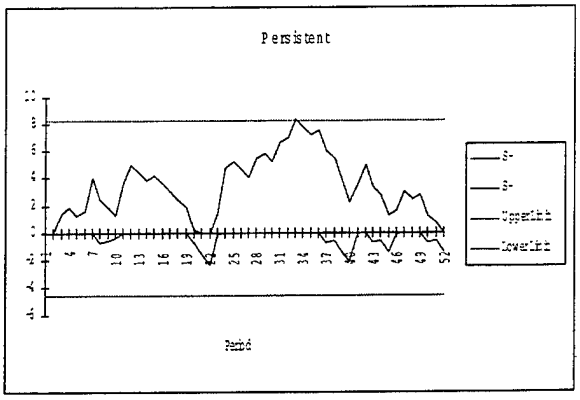


Figure 10: Chart for persistent departures for the NYCPD data. This chart signals an upward shift in the average number of reported rapes in the precinct, beginning at week 23, and signaled at week 33.

monitoring, say, division–level data and Army–level data simultaneously, and seeing if emerging trends were local or Army–wide.

4 Example - DUIs in a division

This military example is based on the number of incidents of Driving Under the Influence (DUI) for an armored division in Europe during the first six months of 1997. We assume that the number of incidents for a given reporting period follows a Poisson distribution, with mean proportional to the number of assigned soldiers and days in the reporting period.

We believe that the true average number of incidents is about 10 per week. We obtained that number by looking at previous historical data, but we could have also obtained it by a formal elicitation of a prior using the methods of Section 2.3. We elicit $\alpha = 20$ and $\beta = 2$, which means that we believe the strength of our prior opinion is the equivalent of two weeks data.

We select an ARL of 500 for the run lengths for both the isolated and persistent departure charts. The average population of the Division is 19,600 soldiers, and the reporting period is seven days. From *ANYGETH.exe*, we obtain the decision intervals (h^+, h^-) and the values of k^+ and k^- for the input screen.

The preceding information is entered in the input screen shown in Figure 18.

4.1 Analysis

The isolated chart signals three unusually bad periods, at weeks 20, 21, and 24. The persistent chart signals an upward increase in the average rate, and estimates that it began at period 19, the last time that S^+ was on the horizontal

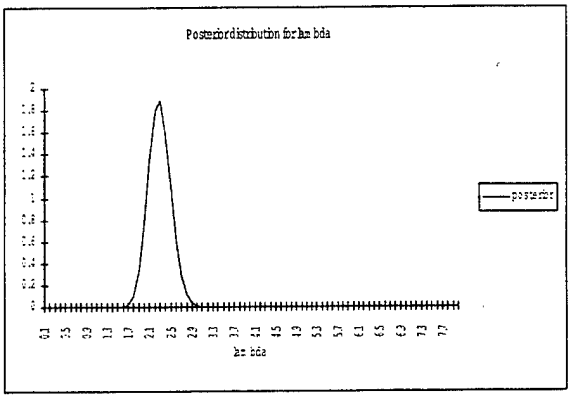


Figure 11: The posterior distribution for λ after all 52 observations of the 75th precinct rape data.

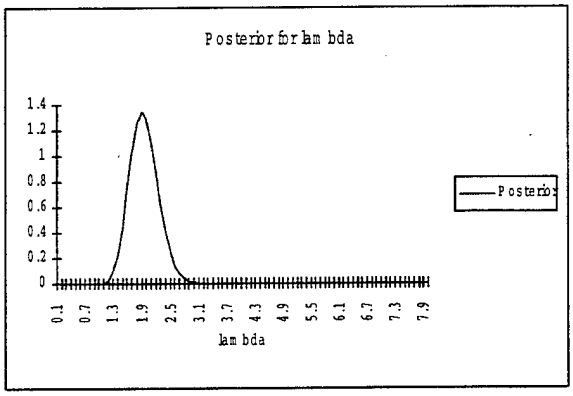


Figure 12: The posterior distribution for λ after the first 22 observations of the 75th precinct rape data.

axis. Clearly the Division has experienced a significant change in DUI behavior, and the leadership should investigate and react immediately.

We know the process has changed, so we investigate why this might have occurred. We are greatly aided in the investigation by knowing about when the change occurred. In this case, this roughly corresponds to the time period when the troops of 1st AD were notified officially that they would be redeployed to Bosnia.

In the meantime, we should restart the charts to monitor the process at its new level. We will want to know if the DUIs remain at this new average level, if they decrease, or if they further increase. The restarted chart will give us that information.

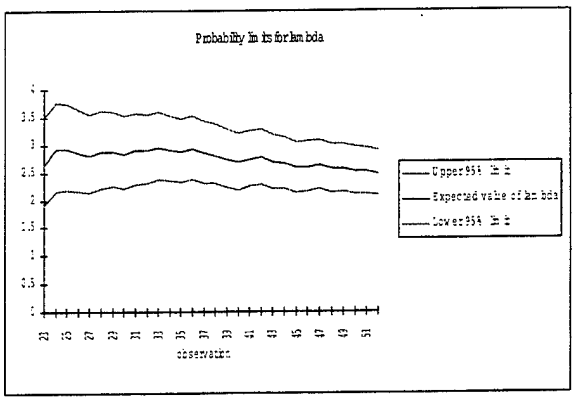


Figure 13: The restarted plot of the posterior mean and 95% probability limits for λ through observation 52 of the 75th precinct rape data, with informed prior.

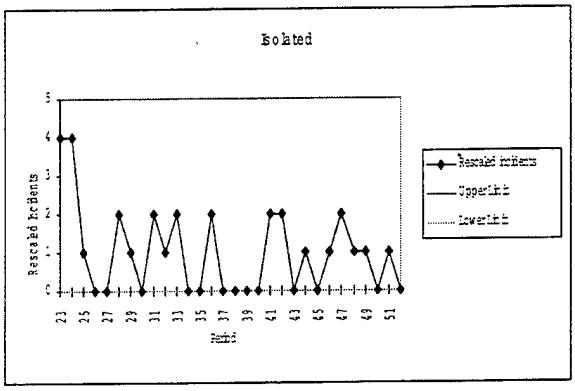


Figure 14: The restarted predictive control chart for an isolated departure through observation 52 of the 75th precinct rape data, with informed prior.

5 Organizational Implications

There is a saying in the Army, "The unit does well those things the boss checks." Setting up a monitoring system using the methods of this paper can have many effects on an organization. They can be good and bad.

Perhaps the best effect is that these tools focus leaders on the distinction between usual and unusual behavior. Over reaction to a process in control can send the process out of control. The methods provide graphical evidence of when extraordinary intervention is and is not warranted. This was illustrated in the discussion of the DoD aviation data.

Second, these tools provide sharp methods for distinguishing between usual and unusual behavior, especially when there are small persistent changes. We saw that the senior leadership of the New York Police Department was unaware

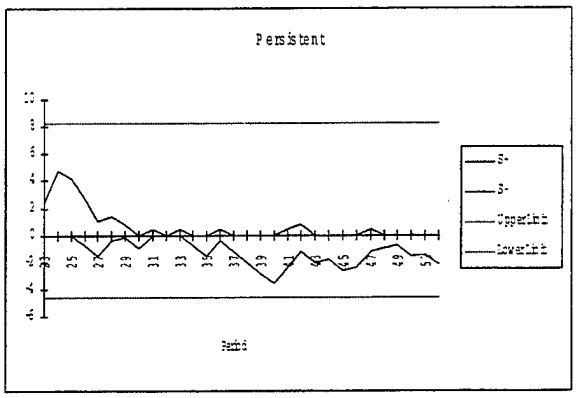


Figure 15: The chart for a persistent change, restarted after week 23 for the 75th precinct data. Note there is no evidence of a further change in the average number of reported rapes.

of the persistent trends which had emerged in the number of reported rapes at precinct and city level in 1996, until these methods were applied to the data.

Third, because the methods are spreadsheet based, they can be maintained easily and require no specialized software. This is a huge advantage, because organizations will be less resistant to using easily implemented methods.

There are some implications which are of concern. We must continue to emphasize that while the level of misconduct or accidents might be at a usual level, no misconduct or accident is ever acceptable. We continue to work vigorously to identify and eliminate the causes of these shortcomings. That is the function of the systems which the organization has in place already to deal with these issues. These tools identify learning opportunities and periods where extraordinary action is warranted. We must be vigilant to avoid sending the message that "usual" is "acceptable." It is not, if soldiers are being hurt. This really requires increased education for Army leaders in the power and limitations of applying statistical thinking to managing the Army.

Secondly, we must be careful not to overburden subordinate units with reporting requirements. There is a trade-off between the frequency of reports and the speed of detection of persistent shifts. However, there is an organizational cost to meeting reporting requirements, especially in the time of subordinate leaders — an already scarce commodity.

Third, we must be sensitive to the effects of data collection. If we continuously ask female soldiers, "Have you been harassed today? Have you been harassed to day?", we can create a hostile working environment for both the questioners and the questioned. If we ask for additional demographic identification, such as race, we can further aggravate soldiers.

Last, we must remind soldiers that the systems that are not being monitored are also important. Not charting may send the message that a process is unimportant, when it might just be unmeasurable.

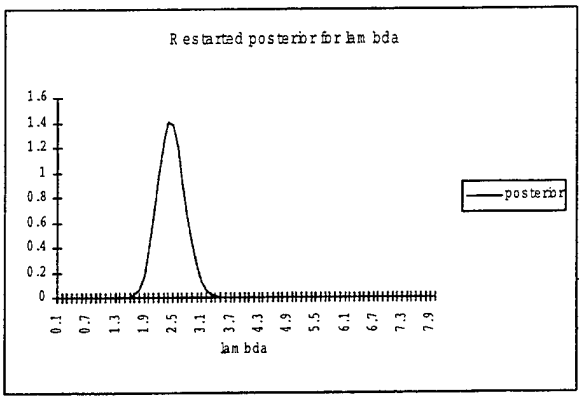


Figure 16: The restarted posterior distribution for λ for weeks 23–52, as of week 52, of the 75th precinct rape data, with informed prior. Note the shift to the right when compared with Figure 12.

6 Acknowledgements

I am grateful to Professor Doug Hawkins of the University of Minnesota, my coauthor of *Cumulative Sum Control Charts and Charting for Quality Improvement*, for his assistance in developing the methods of this paper.

The article was greatly improved as the result of several discussions with LTC Pat Driscoll, Associate Dean for Technology, USMA. I am also grateful to Major Steve Charbonneau, of the Department of Systems Engineering, USMA, for his help with Visual Basic programming for the Excel macros used in this article.

Lieutenant Tonya Cultice worked on these data sets as a cadet during her senior year at USMA. She and her advisor, LTC Greg Dardis of the Department of Behavioral Science and Leadership, contributed to the understanding of the organizational implications of these monitoring schemes.

The data for the DUI example in this paper was provided by the Provost Marshal's Office of the United States Army's 1st Armored Division, Bad Kruez-nach, Federal Republic of Germany, and is gratefully acknowledged.

The data for the aviation mishaps was provide by Mr. John Lemke, Assistant Deputy Undersecretary of Defense for Safety and Occupational Health, and is gratefully acknowledged.

The data on rapes was provided by the Office of the Chief of Department of the New York City Police Department. The cooperation and courtesy of Detective Sergeant John Yohe and Detective Josef Falletta are gratefully acknowledged.

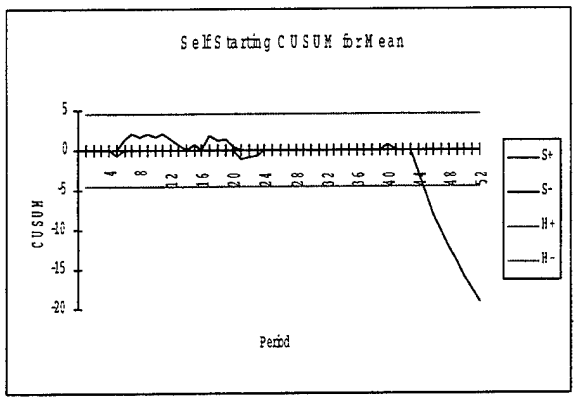


Figure 17: A self–starting control chart for the mean of a normal distribution applied to the NYCPD rape data for the entire city.

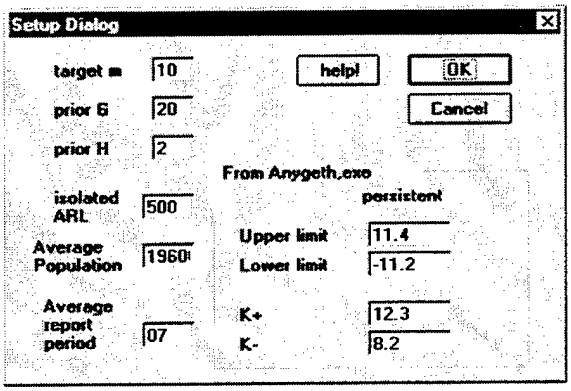


Figure 18: Input screen for Excel program for predictive Poisson control scheme for the Division DUI example.

7 Conclusion

Discrete events well modeled by the Poisson distribution with mean proportional to either time or number of items are everywhere in the Army. This paper has shown a method for monitoring their occurrence to detect unusual periods or persistent shifts in the underlying rate.

These methods have application to misconduct and other social behavior, as well as to weapons systems, vehicles, aircraft, and tools, among others. They offer managers a means for monitoring the performance of these key pieces of equipment, when extensive historical data is missing or when the process does not stay stable for long periods of time.

The unique ability of these methods to learn from the data, and to be restarted when necessary, make them well suited for an ever-changing Army.

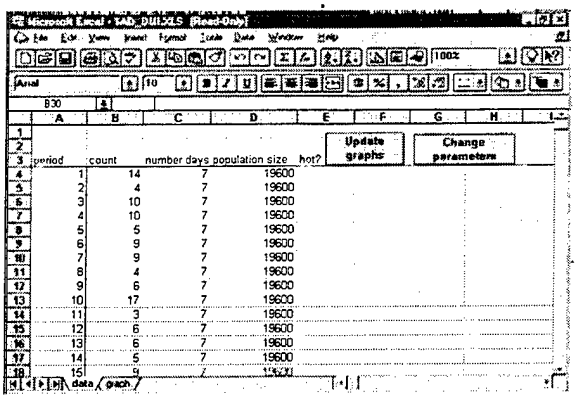


Figure 19: Data screen for Excel program for predictive Poisson control scheme for the Division DUI example.

The Excel–based implementation makes them accessible to soldiers without requiring special software.

These methods are ready for implementation by the Army. They have been implemented into spreadsheet form for ease of use. As of January, 1998, they are being fielded with the operations research analysts supporting the 1st Armored Division in Bosnia. While they require some training in their setup and interpretation, the graphical output is easy to follow. The author hopes that they will soon be part of the leader tool kit for monitoring key command indicators in the entire Army.

8 References

1. Aitchison, J. and I. R. Dunsmore (1975) *Statistical Prediction Analysis.* Cambridge: Cambridge University Press.

2. Bickel, P. J., and Doksum, K. A. (1977) *Mathematical Statistics: Basic Ideas and selected topics.* San Francisco: Holden-Day.

3. Geisser, Seymour (1993) *Predictive Inference: An Introduction* New York: Chapman & Hall.

4. Hawkins, Douglas M. (1987) Self-starting cusums for location and scale. *The Statistician* Vol. 36. pp. 299-315.

5. Hawkins, D. M. and Olwell, D. H. (1998) *Cumulative Sum Control Charts and Charting for Quality Improvement.* New York: Springer-Verlag.

6. Montgomery, Douglas C. (1991) *Introduction to Statistical Process Control.* 2nd Ed. New York: Wiley.

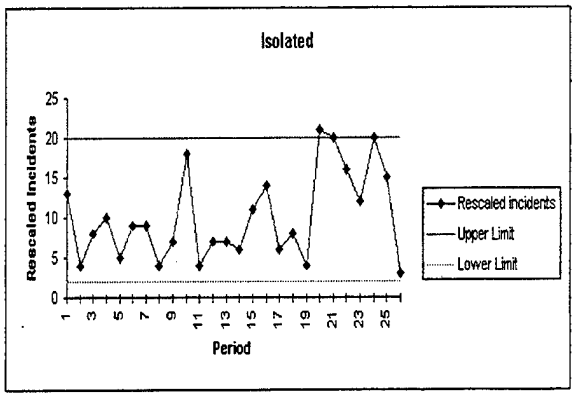


Figure 20: Chart for isolated departures, from Excel program for predictive Poisson control scheme for DUI example. This chart signals weeks 20, 21, and 24 as unusually high periods.

7. Moustakides, G. V., (1986), "Optimal stopping times for detecting changes in distributions," *Annals of Statistics*, 14, 1379–1387.

8. Olwell, David H. (1996) *Topics in Statistical Process Control.* Ann Arbor: University Microfilms.

9. Olwell, David H. (1997a) Predictive Quality Control Charts. *Proceedings of the Second Annual US Army Conference on Applied Statistics, 23-25 October 1996.* Ed. Barry Bodt. Aberdeen, MD: US Army Research Lab. Pp. 67–84.

10. Olwell, David H. (1997b) Statistical Process Control of Low Frequency Events. *Proceedings of the 5th Annual US Army Research Laboratory . United States Military Acedmy Technical symposium.* Ed. Don Engen. West Point, NY: US Military Academy. Pp. 115-130.

11. Olwell, David H. (1998) Managing Misconduct: Statistical Process Control applied to Sexual Harassment and Other Crimes in the Military. *1997 Proceedings of the Section on Quality and Productivity of the American Statistical Association.* Fairfax,VA: American Statistical Association.

12. Page, E. S., (1954), "Continuous inspection schemes," *Biometrika*, Vol. 41, 100–115.

13. Page, E. S., (1961), "Cumulative sum charts," *Technometrics*, Vol. 3, pp. 1–9.

14. Perry, J. N. and R. Mead (1979) On the power of the index test to detect spatial pattern. *Biometrics.* Vol. 35, pp. 613–622.

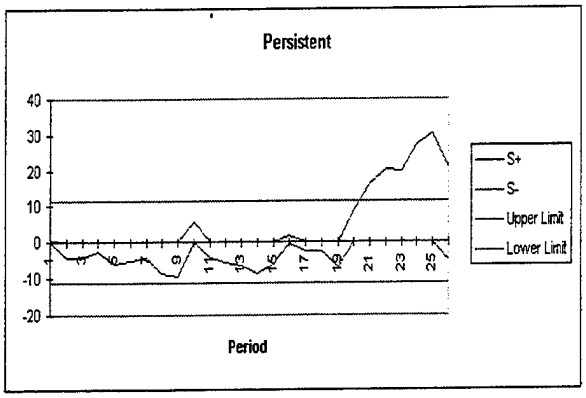


Figure 21: Chart for persistent departures, from Excel program for predictive Poisson control scheme for DUI example. This chart signals a persistent increase in the average number of incidents, which began at week 19 and was signaled at week 21.

15. Quesenberry, C. P., (1991), "SPC Q charts a Poisson Parameter λ: short or long runs," *Journal of Quality Technology* Vol. 23, No. 4, pp. 296–303.

16. Quesenberry, C. P., (1995), "On Properties of Poisson Q Charts for Attributes," *Journal of Quality Technology*, Vol. 27, No. 4, pp. 293–303.

17. Shewhart, Walter A. (1931) *Economic Control of Quality of Manufactured Product.* New York: Van Nostrand.

18. Tierney, Luke (1990) *LISP-STAT*, New York: Wiley.

19. Western Electric (1956) *Statistical Quality Control Handbook*, Indianapolis, IN: Western Electric Corporation.

Using Pearson and Spearman Statistics to Look for a Dependence Relationship between Two Variables

Dr. Lorrie L. Hoffman, Department of Statistics, University of Central Florida
Mr. Dan Corson, IMAGE intern from Florida A&M University

Introduction

The purpose of this study is to make data analysts acutely aware of the differences between two popular statistics: Spearman's rank correlation coefficient and Pearson's chi-square test for independence. Both of these statistics are used to reveal relationships between two variables x and y. Spearman's will uncover linear relationships, i.e. where y can be adequately predicted as a linear function of x. Pearson's chi-square is less stringent in that the relationship need not be linear. The simulations conducted and reported herein will reinforce the fact that Pearson's test can reveal a dependence between x and y while Spearman's will not. This occurs simply because the relationship is not a linear one.

Both statistics are considered nonparametric since knowledge of the underlying original distribution of the data is not required in order to know their sampling distributions. See Practical Nonparametric Statistics, 2nd edition, W.J. Conover, John Wiley and Sons, New York, 1980, for a reference. The exact sampling distribution for Spearman's rank correlation coefficient in the case where x and y are continuous can be derived theoretically via a combinatorial argument (Conover, p. 252). For large sample sizes a normal approximation suffices. The exact distribution for Pearson's test can also be generated by viewing the contingency table (the frequency of occurrences as defined in a two-way classification table) as a realization from a multinomial distribution and considering all possible outcomes. For adequate sized cell counts in the classification (contigency) table a chi-square approximation suffices (Conover, p. 160).

Recall that Spearman searches for relationships by relying on the rank of each x among all x's and of y among all y's and testing whether large x ranks pair with large y ranks (positive correlation) or large x ranks pair with small y ranks (negative correlation). Pearson's approach can be explained as one which slices up the 3-D histogram (the cell counts being the heights corresponding to the varying x columns and y rows) and then checks to see if those slices (near equivalent to conditional distributions) are proportionally identical.

Methodology

We needed to choose a family of distributions which were capable of exhibiting non-linear dependence. A mixture of bivariate normals meets this criteria. We have restricted our attention to

$$f(x,y) = .5/\sqrt{2*\pi*(1-\rho_1^2)}*\exp((-.5/(1-\rho_1^2))*(x^2-2*\rho_1*x*y+y^2))$$
$$+.5/\sqrt{2*\pi*(1-\rho_2^2)}*\exp((-.5/(1-\rho_2^2))*(x^2-2*\rho_2*x*y+y^2)) \qquad (1)$$

$$-\infty < x < +\infty \quad \text{and} \quad -\infty < y < +\infty.$$

That is, it is equally likely we sample an <x,y> pair from a normal with a zero-vector mean, variances of 1 and a covariance of rho(1) as to sample from a normal with zero-vector mean, variances of 1 and a covariance of rho(2).

Using the suggested generation scheme in Multivariate Simulation, M.E. Johnson, John Wiley and Sons, New York, 1985, and SPSS software version 6.1 for Windows we generated 300 pairs for each investigation. We restricted ourselves to five cases consisting of setting rho(1) to .50 and rho(2) = -.75, -.50, .0, .50, .75 . We looked at k=10 runs of 300 pairs for each of the five cases.

The required steps to generate this mixed bivariate normals are:

1. Generate U, where U is from a uniform distribution on (0,1)
2. Generate X11 from a normal distribution with mean of 0 and variance of 1
3. Generate X12 from a normal distribution with a mean of rho(1)*X11 and a variance of (1-rho(1)**2)
4. Generate X21 from a normal distribution with mean of 0 and variance of 1
5. Generate X22 from a normal distribution with a mean of rho(2)*X21 and a variance of (1-rho(2)**2)
6. X=X11 if U<=1/2
   Y=X12 if U<=1/2
   X=X21 if U>1/2
   Y=X22 if U>1/2.

The Spearman test statistic value is calculated in SPSS by using the BIVARIATE CORRELATION macro under the STATISTICS drop-down menu and selecting the Spearman option. The Pearson test statistic value is calculated by using the SUMMARIZE-DESCRIPTIVE-CROSSTABS macro under the STATISTICS drop-down menu after classifying the x and y values into these categories: Category is -2 if x< -1.0, category is -1 if -1.0<= x < 0.0, category 1 if 0.0<= x < 1.0 and category 2 if x>=1.0 with an identical scheme for y. This creates a 2 by 2 contingency table with 16 cells for generating Pearson's chi-square test statistic. Note that this gridding decision is arbitrary and very well may affect the ability of Pearson's test statistic to discern independence. This concern is left to later research.

The p-values reported herein are gleaned from the SPSS output which relies on the two-sided tail areas under the approximate normal distribution for Spearman and on a chi-square approximation for Pearson.

Since we know the true original distribution (mixed bivariate normals) and have derived the conditional distributions and correlation associated with each then these simulations are conducted in order to assess the performance of these test statistics, both individually and in relationship to each other. (Note: here n=300. We could have varied this but believe it mimics studies we often see our colleagues in Education and Business conduct.) We run k=10 simulations for each of the five cases above.

It is direct to show that the conditional distributions associated with f(x,y) are:

$$f(y|x) = .5 / \sqrt{2 * \pi * (1 - \rho_1^2)} * \exp((-.5 / (1 - \rho_1^2)) * (y - \rho_1 * x)^2))$$

$$+ .5 / \sqrt{2 * \pi * (1 - \rho_2^2)} * \exp((-.5 / (1 - \rho_2^2)) * (y - \rho_2 * x)^2)$$

$$- \infty < y < + \infty \quad . \quad (2)$$

And the correlation of x and y is

$$.5 * \rho_1 + .5 * \rho_2 . \qquad (3)$$

Recall that if the conditional distributions are not identical for each X then the definition of independence (P(Y=y|X=x) = P(Y=y)) is violated.

In none of these cases is independence true and in only one case where rho(1)=.5 and rho(2) = -.5 is the correlation zero. This statement is reinforced below in the *Results* section by presenting the graphs of the conditional distribution for x= 0 and for x=1 for each of the five cases, and by calculating the correlation of x and y. This means that both Spearman and Pearson should reject their respective hypotheses of H0: rho = 0 (except for the case where rho(1)=.5 and rho(2)= -.5) and reject H0: x and y are independent.

Results

The results will be presented case by case and then discussed collectively at the end. We summarize the rejection/acceptance activity in Table 1 and present the calculated correlation values in Table 2. For each case we show the following: 1) a 3-D wireframe plot and 2) a contour plot (both graphics generated by MINITAB for WINDOWS) plus 3) graphs of the conditional distribution functions when x = 0 (dashed line) and x = 1 (solid line) via SPSS plots. Table 1 reports the p-values from the k=10 runs for Spearman and reports the p-values from the k=10 runs for Pearson for the cases where rho(2) is -.75, -.50, and .0 (omitting the rho(2) = .50 and .75 since the p-values for these other two cases were all less than .0005). Included at the bottom of Table 1 for both Spearman and Pearson are the number of rejections of the H0 hypotheses which occur using an alpha level of .10. Note that the smallest expected cell count for all of the contingency tables was 4.3 and so the chi-square approximation should be adequate. Table 2 contains the Spearman correlation coefficients calculated from the 300 points for each of the k=10 runs for every one of the five cases.

For the case where rho(1)=.50 and rho(2)= -.75 the 3-D plot presents a shape which has a lop-sided four-leaf clover look to its cross-section slices (Figure 1). This is best

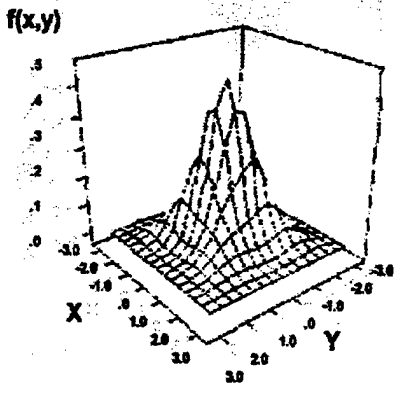


Figure 1. 3-D plot with rho(2)= -.75

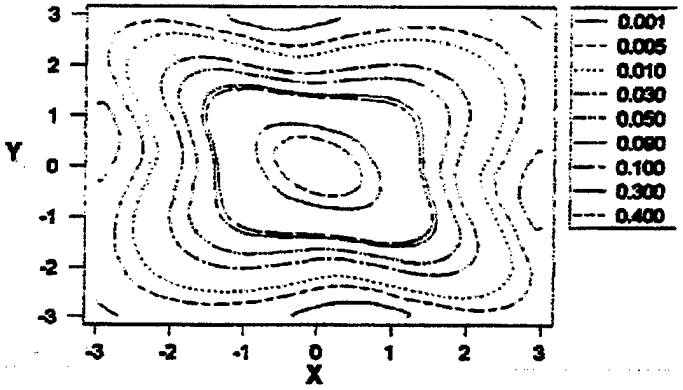


Figure 2. Contour plot with rho(2)= -.75

understood by viewing the contour plots (Figure 2). This realization is an obvious result of combining an equal mix of bivariate normals: one with its true regression line running through (0,0) and (1, 0.5) and the other one through (0,0) and (1,-.75). The dependence between x and y is evident when one views two arbitrarily selected conditional distributions: one where x=1 and the other when x=0. These are presented in Figure 3. The graphical evidence should lead us to expect that both Pearson's chi-square statistic (by noting the dissimilarity of the conditional distributions) and Spearman's correlation coefficient (by noting the slight domination of a downward sloping tendency in the contour plots) will uncover a dependency between x and y. Those values are presented in Tables 1 and 2 below.

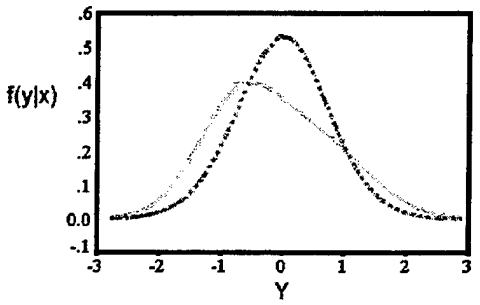


Figure 3. Conditional distributions at x=0 (dashed) and at x=1 (solid) rho(2)= -.75

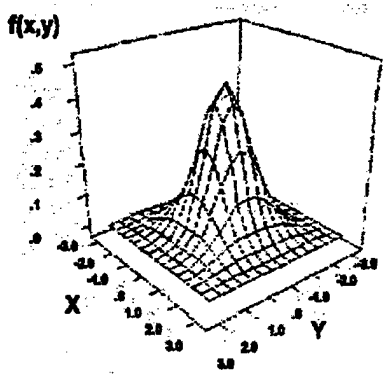


Figure 4. 3-D plot with rho(2)= -.5

For the case where rho(1)=.50 and rho(2)= -.50 the 3-D plot presents a shape which has a symmetric four-leaf clover look to its cross-section slices (Figure 4). The difference when comparing to the previous case is reinforced by viewing the contour plots (Figure 5). This graphical portrayal is an obvious result of combining an equal mix of bivariate normals: one with its true regression line running through (0,0) and (1, 0.5) and the other one through (0,0) and (1,-.5). The dependence between x and y is evident when one views two arbitrarily selected conditional distributions: one where x=1 and the other when x=0. These are presented in Figure 6. In this case the conditional distributions are symmetric. This is easily confirmed by showing f(y|X=x) = f(y|X=-x) in equation 2. The graphical evidence should lead us to expect that Pearson's chi-square statistic will uncover a dependency between x and y. Since there appears to be no propensity for the values to cluster about any line we would expect Pearson's correlation coefficient to be near zero. Those test statistical values are presented in Tables 1 and 2 below.

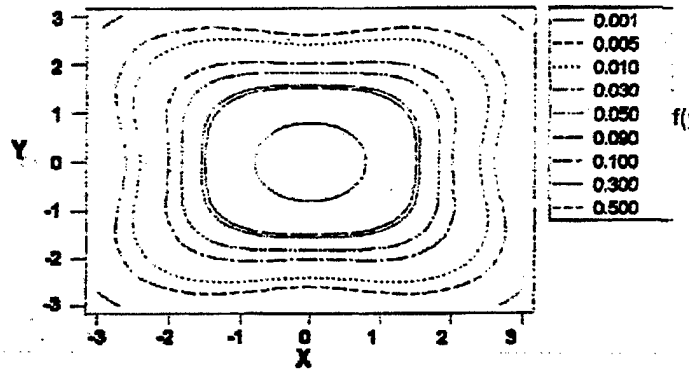


Figure 5. Contour plot with rho(2)= -.50

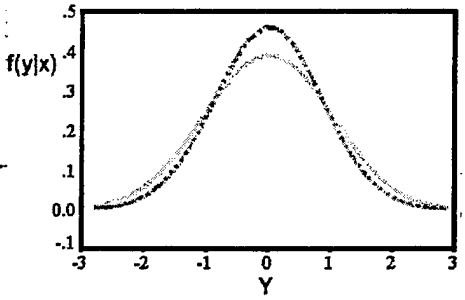


Figure 6. Conditional distributions at x=0 (dashed) and at x=1 (solid); rho(2)= -.50

For the case where rho(1)=.50 and rho(2)= .0 the 3-D plot presents a shape which has an elliptical look to its cross-section slices (Figure 7). This difference from previous shapes is best seen in the contour plots (Figure 8). This graphical image occurs due to combining an equal mix of bivariate normals with circular cross-sections (the one with zero correlation) and the other bivariate normal with its regression line running through (0,0) and (1, .5).

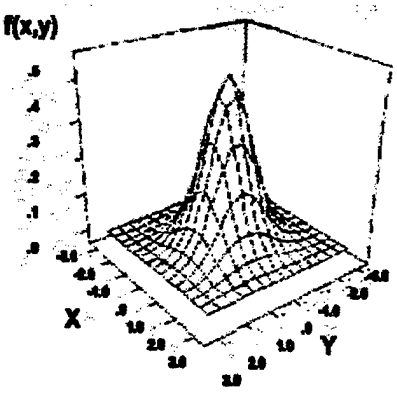


Figure 7. 3-D plot with rho(2)= .0

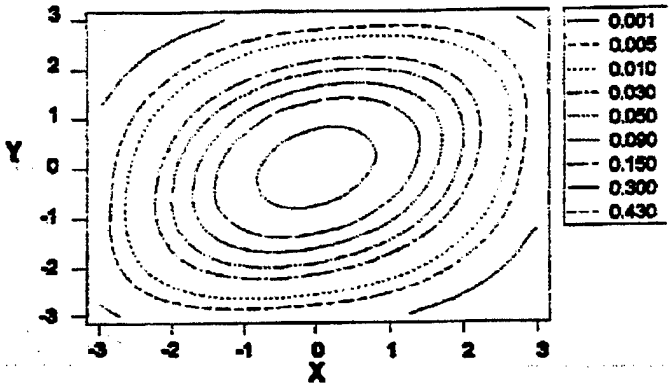


Figure 8. Contour plot with rho(2)= .0

The dependence between x and y is evident when one views those two arbitrarily selected conditional distributions. These are presented in Figure 9. In this case the conditional distributions exhibit dissimilarities: at x=1 there is more chance of observing positive y-values (note the larger area to the right of y=0 than to the left) whereas when x=0 it appears equally likely one would observe a positive or a negative y-value. The graphical evidence points to a significant Pearson's chi-square statistic. Since there appears to be clustering about a line with positive slope then we would expect Pearson's correlation coefficient to be positive. Those test statistical values are presented in Tables 1 and 2 below.

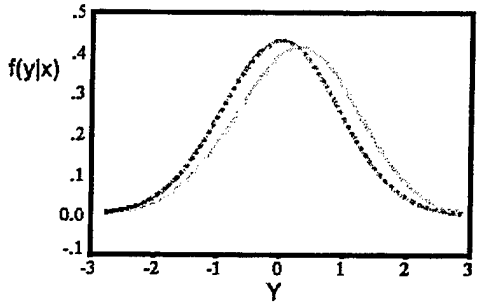


Figure 9. Conditional distributions at x=0 (dashed) and at x=1 (solid) rho(2)= .0

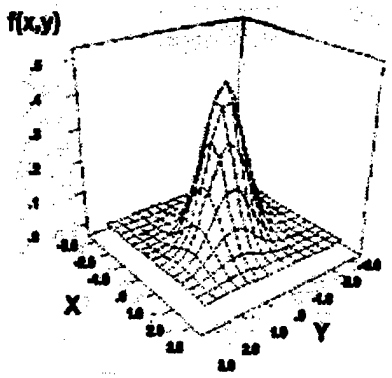


Figure 10. 3-D plot with rho(2)= .50

For the case where rho(1)=.50 and rho(2)= .50 the 3-D plot is merely the bivariate normal with a zero mean vector, variances of 1, and a correlation coefficient equal to .50 (Figure 10). The distinction of previous shapes may now be best understood by comparing them to this 'non-mixed' bivariate normal (Figure 11).

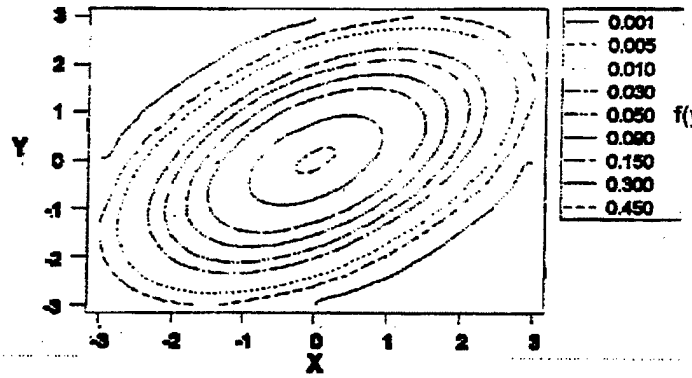


Figure 11. Contour plot with rho(2)= .50

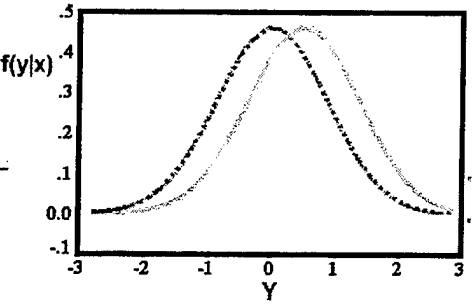


Figure 12. Conditional distributions at x=0 (dashed) and at x=1 (solid); rho(2)= .50

The dependence between x and y is evident when one views those two arbitrarily selected conditional distributions. It is also well understood theoretically since we know that the conditional distributions of a bivariate normal are also normal with a mean and variance equal to, respectively, *(4)*

$$\mu_y + \rho^*(\sigma_y / \sigma_x)^*(x - \mu_x) \quad , \sigma_y(1 - \rho^2)$$

So in this case when x=0 Figure 12 presents a normal with a mean of 0 and variance of .75 and when x=1 a normal with mean of .50 and variance of .75. Since the distributions are not identical we conclude that there is a dependence between x and y and should calculate a Pearson's chi-square statistic which is significant. Visually we see a clustering about a line with positive slope and thus expect Pearson's correlation coefficient to be positive and can be found in Table 2 below.

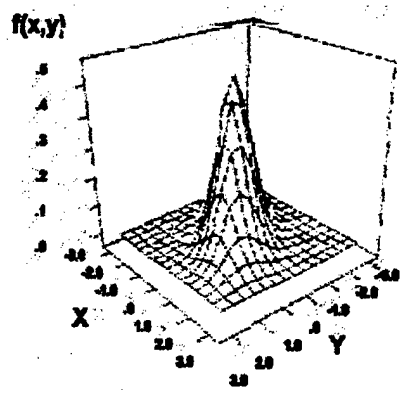


Figure 13. 3-D plot with rho(2)= .75

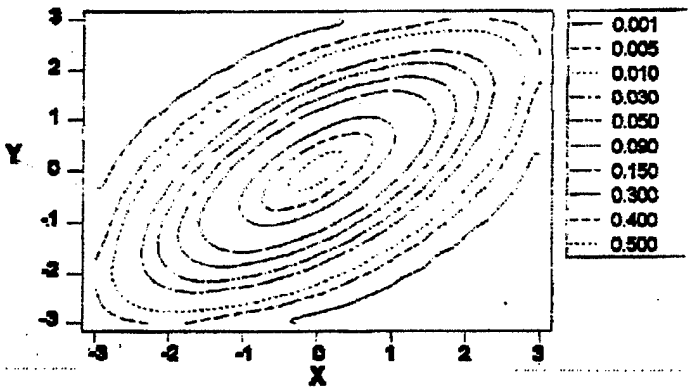


Figure 14. Contour plot with rho(2)= .75

For the last case where rho(1)=.50 and rho(2)= .75 the 3-D plot appears extremely narrow (Figure 13). This concentrated clustering about a regression line with positive slope is quite evident in Figure 14.

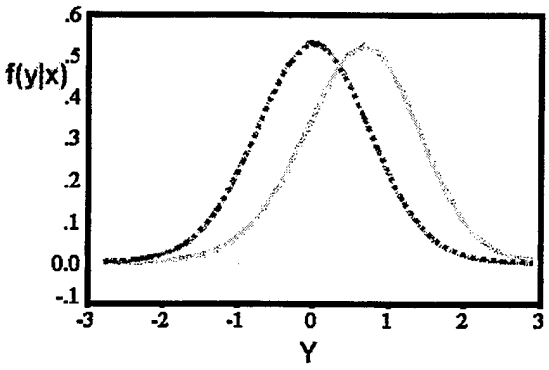


Figure 15. Conditional distributions at x=0 (dashed) and at x=1 (solid) ; rho(2)= .75

The dependence between x and y is showcased by viewing those two arbitrarily selected conditional distributions. These are presented in Figure 15. In this case the conditional distributions exhibit dissimilarities: at x=1 there is a very large chance of observing positive y-values (note the larger area to the right of y=0 than to the left) whereas when x=0 it appears equally likely one would observe a positive or a negative y-value. A significant Pearson's chi-square statistic should result. Since there appears to be clustering about a line with positive slope then we would expect Pearson's correlation coefficient to be positive (Table 2).

rho(2)	-.75		-.50		.0	
run	Spear	Pears	Spear	Pears	Spear	Pears
1	0.059	0.003	0.989	0.032	0.000	0.015
2	0.001	0.002	0.935	0.024	0.004	0.126
3	0.001	0.000	0.255	0.884	0.003	0.000
4	0.006	0.000	0.185	0.001	0.000	0.000
5	0.210	0.000	0.298	0.043	0.007	0.075
6	0.053	0.003	0.487	0.023	0.000	0.008
7	0.264	0.000	0.858	0.067	0.000	0.005
8	0.000	0.000	0.293	0.026	0.001	0.145
9	0.059	0.003	0.128	0.051	0.000	0.000
10	0.062	0.005	0.495	0.441	0.043	0.000
#rejct	8	10	0	8	10	8

Table 1. Pearson and Spearman p-values when rho(1)=.5 and rho(2) varies

rho(2)	-0.750	-0.500	0.000	0.500	0.750
run					
1	-0.109	-0.001	0.221	0.476	0.608
2	-0.193	-0.005	0.167	0.416	0.570
3	-0.197	-0.066	0.177	0.439	0.583
4	-0.157	-0.077	0.264	0.414	0.561
5	-0.073	-0.060	0.156	0.496	0.629
6	-0.112	0.040	0.225	0.466	0.593
7	-0.065	-0.010	0.301	0.529	0.654
8	-0.223	0.060	0.193	0.474	0.612
9	-0.109	-0.088	0.321	0.575	0.681
10	-0.108	0.039	0.199	0.416	0.593
ave	-0.135	-0.017	0.222	0.470	0.608
rho	-0.125	0.000	0.250	0.500	0.625

Table 2. Calculated rho values with theoretical rho value (at bottom)

Conclusions

In each of these cases involving mixtures of bivariate normals Spearman's rho does a fine job of estimating the true underlying correlation coefficient as exhbited in Table 2. This translates to excellence in testing where we see in Table 1 that Spearman mostly makes the correct decision on whether to accept or reject H0: rho = 0. Pearson does equally as well. Further investigations should delve into the particulars relating to the power of these tests. That was not our purpose here.

Our purpose is to spotlight the data in Table 1 in the center columns associated with the mixture of bivariate normals with a correlation coefficient of zero (i.e. when rho(1)= .5 and rho(2)= - .5). Spearman concludes correctly that rho is zero (in 10 of 10 runs at alpha of .10) and Pearson concludes correctly that x and y are dependent (in 8 of 10 runs at alpha of .10). Note that based on Spearman some data analysts might erroneously report that x and y are 'independent' rather than correctly saying 'uncorrelated' as is true in this instance. This false deduction occurs due to an error in thinking. A data analyst may forget that linear dependence is not the only kind of dependence. In the other four cases we present here either Pearson or Spearman statistics adequately reflect the dependence characteristic inherent in the data, but in the case of our mixed bivariate normals with rho(1)= .5 and rho(2)= -.5 a researcher must pursue an avenue of investigation beyond merely asking about linear correlation in order to reveal a dependence relationship. This cautionary note was the impetus for this work of finding ready examples of non-linear dependence and the subsequent characterization of those distributions.

THE JOINT DISTRIBUTION OF THE MEAN AND AN EXTREMUM OF A NORMAL SAMPLE, WITH APPLICATIONS TO QUALITY CONTROL

Mark G. Vangel
National Institute of Standards and Technology
Statistical Engineering Division
Building 820, Room 353
820 West Diamond Avenue
Gaithersburg, MD 20899-0001

Abstract

In some industrial applications one compares a sample mean and minimum, or a mean and maximum, to reference values, and determines if the lot from which the sample was taken is acceptable, or if further investigation of this lot is indicated. Because the exact joint distribution of an extremum and the mean of a normal sample is complicated, establishing these reference values using statistical considerations typically involves crude approximations or simulation, even under the assumption of normality. The purpose of this article is to use the saddlepoint method to develop a fairly simple and very accurate approximation to the joint cdf of the mean and an extremum of a normal sample. Tables for use in establishing acceptance criteria are also provided, and the use of these tables is illustrated with an example.

KEY WORDS: saddlepoint, control chart, acceptance testing

INTRODUCTION

Let $Y_i, i = 1, \ldots, n$ be an *iid* sample from a normal population, and denote the sample order statistics as $Y_{(1)} < Y_{(2)} < \ldots < Y_{(n)}$. In some industrial applications, one compares $Y_{(1)}$ and $\bar{Y}$, or $Y_{(n)}$ and $\bar{Y}$, to reference values, to determine if the lot from which the sample was taken is acceptable, or if this lot should be investigated further.

[1] The results of this article are potentially applicable in diverse industries. For example, emphasis has long been placed on checking sample means and minima of lots of various packaged goods (Brickenkamp et al., 1988; Croarkin and Yang, 1987). Also, the sample means and minima are used in the testing of batches of raw material by many manufacturers of composite materials. And means and maxima of power loss of sampled motors have been proposed for use in testing whether manufactured motors comply with labeled motor efficiencies (Stricklett and Vangel, 1997).

If these reference values are established using statistical considerations, then it is likely that either crude approximations or simulation will be employed, because the exact joint distribution of an extremum and the mean of a normal sample is complicated, with complexity increasing rapidly with n (Murty and Bissinger, 1982). The purpose of this article is to develop an approximation for this joint cdf based on the saddlepoint method (see, e.g., Barndorff-Nielsen and

[1] Approved for public release; distribution is unlimited.

Cox, 1989, Section 4.3), which is very accurate and straightforward to compute. The use of this distribution in acceptance testing is illustrated with an example.

AN APPROXIMATE JOINT CDF fOR $Y_{(1)}$ AND $\bar{Y}$

Assume that the random variables Y_i are *iid* normal with mean μ and variance σ^2. Since we will assume that μ and σ can be regarded as approximately known from previous data, we can, without loss of generality, employ the standardized sample $X_i \equiv (Y_i - \mu)/\sigma$, having order statistics $X_{(i)} \equiv (Y_{(i)} - \mu)/\sigma$.

Let $F_{X_{(1)}}(t)$ denote the cumulative distribution of $X_{(1)}$, and let $F_{\bar{X}}(t)$ be the cdf of $\bar{X}$; that is

$$F_{X_{(1)}}(t) = \Pr(X_{(1)} \leq t) = 1 - (1 - \Phi(t))^n,$$

and

$$F_{\bar{X}}(t) = \Pr(\bar{X} \leq t) = \Phi\left(\sqrt{n}t\right),$$

where $\Phi(\cdot)$ is the standard normal cdf.

The primary objective of this paper is to develop and illustrate an approximation to

$$F_{X_{(1)},\bar{X}}(t_1, t_2) = \Pr(X_{(1)} \leq t_1 \text{ and } \bar{X} \leq t_2).$$

Let $\phi(t)$ denote the normal density, and let

$$h(t) = \frac{\phi(t)}{1 - \Phi(t)}$$

be the normal hazard function. In the Appendix, we derive the saddlepoint approximation $\tilde{F}_{X_{(1)},\bar{X}}(t_1, t_2)$, for $t_1 \leq t_2$, where

$$\tilde{F}_{X_{(1)},\bar{X}}(t_1, t_2) = \frac{\int_{-\infty}^{t_*} \Phi(\sqrt{n}t_2) A(t)dt + \int_{t_*}^{\infty} \Phi\left\{\sqrt{n}\left[t_1 + \frac{n-1}{n}(h(t) - t)\right]\right\} A(t)dt}{\int_{-\infty}^{\infty} A(t)dt}, \qquad (1)$$

$$\begin{aligned} A(t) &= h^{-(n-1)}(t) \exp\left[\frac{(n-1)^2}{2n}(h(t) - t)^2 + (n-1)t(h(t) - t)\right] \\ &\quad \cdot\sqrt{1 - h^2(t) + th(t)}, \end{aligned}$$

and where t_* is the (unique) solution to the equation

$$\frac{n-1}{n}(h(t_*) - t_*) = t_2 - t_1. \qquad (2)$$

The approximate cumulative distribution (1) satisfies the conditions of a bivariate cdf. The derivation appears in an appendix because it is fairly technical; however, the details are elementary, and somewhat interesting.

A contour plot of $\tilde{F}_{X_{(1)},\bar{X}}(t_1, t_2)$ for sample size $n = 5$ is displayed in Figure 1. Note the rather sharp 'corners' on the contours: if $t_2 >> t_1$, then

$$F_{X_{(1)},\bar{X}}(t_1, t_2) \approx F_{X_{(1)},\bar{X}}(t_1, \infty) = F_{X_{(1)}}(t_1),$$

and if $t_1 << t_2$,

$$F_{X_{(1)},\bar{X}}(t_1, t_2) \approx F_{X_{(1)},\bar{X}}(-\infty, t_2) = F_{\bar{X}}(t_2).$$

When choosing reference values to be used for acceptance testing, one would typically try to make essential use of both the mean and the extremum; this corresponds to being near the corners

of the contours in Figure 1. One reasonable condition to impose is that, if $F_{X_{(1)},\bar{X}}(t_1,t_2)=\alpha$, then

$$F_{X_{(1)}}(t_1)=F_{\bar{X}}(t_2)=\alpha/2. \tag{3}$$

This ensures that if a future lot comes from the same population as the lots used to establish the acceptance criteria, then

$$\Pr\left(Y_{(1)}\le\mu+t_1\sigma\right)=\Pr\left(\bar{Y}\le\mu+t_2\sigma\right).$$

The (t_1,t_2) pairs for which this is true are given by the equation

$$t_2=\Phi^{-1}\left[1-(1-\Phi(t_1))^n\right]/\sqrt{n}. \tag{4}$$

For $n=5$, (4) is displayed as the broken curve in Figure 1.

THE ACCURACY OF THE APPROXIMATION

The saddlepoint approximation used in the derivation of (1) is asymptotic in n, and hence can be expected to be accurate for large n. However, the saddlepoint method often leads to approximations which are reasonably accurate for *all* sample sizes, and this turns out to be the case for $\tilde{F}_{X_{(1)},\bar{X}}(t_1,t_2)$. As an example, if $n=3$, $t_1=-2.5$, and $t_2=-1.20404$ (which is the value given by (4)), then from a 1,000,000-replicate simulation, we have that, $F_{X_{(1)},\bar{X}}(-2.5,-1.20404)\approx .005421$ with an approximate 95% (binomial) confidence interval of

$$.005274\le F_{X_{(1)},\bar{X}}(-2.5,-1.20404)\le .005567.$$

For this example, $\tilde{P}(-2.5,-1.20404)=.005226$ is only slightly below the simulation uncertainty interval.

Rather than performing a large simulation, we have relied on checking special cases (such as the above example), and on comparison with the exact result for $n=2$. If $n=2$, it is easy to show that

$$F_{X_{(1)},\bar{X}}(t_1,t_2)=2\int_{-\infty}^{t_1}\Phi(2t_2-x)(1-\Phi(x))\phi(x)dx. \tag{5}$$

Contours of (5) are displayed in Figure 2, along with the corresponding contours of (1). Note that the approximation is quite accurate, even for this 'worst case' sample size.

TABLES

We provide tables for the following situations.

1. One intends to reject a lot if either $Y_{(1)}\le\mu-k_1\sigma$ or $\bar{Y}\le\mu-k_2\sigma$; or else
2. one plans to reject if either $Y_{(n)}\ge\mu+k_1\sigma$ or $\bar{Y}\ge\mu+k_2\sigma$.

We impose the condition that if an error of the first kind is made, then it is equally likely that we are required to reject because of a sample extremum as it is that we must fail because of the sample mean. That is, calculations for the tables employ the constraint (4).

Tables 1 and 2 provide values k_1 and k_2 such that

$$Q(k_1,k_2)\equiv F_{X_{(1)}}(-k_1)+F_{\bar{X}}(-k_2)-\tilde{F}_{X_{(1)},\bar{X}}(-k_1,-k_2)=\alpha, \tag{6}$$

and

$$F_{X_{(1)}}(-k_1)=F_{\bar{X}}(-k_2)=\alpha/2,$$

where the probability (α) and sample size (n) correspond, respectively, to rows and columns in the tables. The tabulated function $Q(k_1,k_2)$ has the property that, within the accuracy of the saddlepoint approximation,

$$Q(k_1,k_2)\approx\Pr\left(Y_{(n)}\ge\mu+k_1\sigma\text{ or }\bar{Y}\ge\mu+k_2\sigma\right)=\Pr\left(Y_{(1)}\le\mu-k_1\sigma\text{ or }\bar{Y}\le\mu-k_2\sigma\right).$$

POWER

To illustrate the power of an acceptance test based on the mean and an extremum, we let $n = 5$, $\alpha = .001$, and we require that (3) holds under the null hypothesis. Denote the parameters of the normal distribution under the null hypothesis by (μ_0, σ_0), and under the alternative by (μ, σ). Using the function Q defined in (6), it is easy to show that the power of this test is

$$\pi_1(a,b) = \Pr\left[Y_{(1)} \le \mu_0 - k_1\sigma_0, \bar{Y} \le \mu_0 - k_2\sigma_0\right] = Q\left[-b(a+k_1), -b(a+k_2)\right], \tag{7}$$

where $a = (\mu - \mu_0)/\sigma_0$, $b = \sigma_0/\sigma$, $k_1 = 3.7132$, and $k_2 = 1.4687$. We compare this function with the power of a test based on the mean alone,

$$\begin{aligned} \pi_2(a,b) &= \Pr(\bar{Y} \le \mu - z_\alpha\sigma/\sqrt{n}) \\ &= \Phi\left[-b(a\sqrt{n} + z_\alpha)\right], \end{aligned} \tag{8}$$

where $z_\alpha \doteq 3.0902$ is a normal quantile.

Contours of the power differences

$$\delta(a,b) \equiv \pi_1(a,b) - \pi_2(a,b)$$

are displayed in Figure 3. The only situation where the test based on the mean has appreciably more power than that which used both mean and an extremum is when $a \approx -k_2$, and $b > 1$. One can see from (7) that when $a = -k_2$, then $\pi_1(a,b)$ depends on the data only through $Y_{(1)}$. On the other hand, when b is substantially less than 1, $\delta(a,b)$ can be large. So the use of the mean and an extremum is preferable to the use of the mean alone if, in addition to shifts in the mean, one is concerned about detecting increases in the standard deviation. (The use of an extremum also provides obvious protection against spurious extreme observations, but this more a matter of robustness than power.)

An Example

As an example of how the results of this paper can be used in practice, we consider the problem of testing incoming batches of raw material. Aerospace composite materials are often fabricated from large rolls of resin-impregnated graphite fiber. In order to determine whether a newly-arrived roll is a cause for concern, specimens are made, and various mechanical and chemical properties are tested. For strength properties, it has become common in this industry to require that both the average and minimum of the test data exceed certain critical values. Usually these threshold values are arrived at in a somewhat ad-hoc manner. We suggest here an alternative.

Figure 4 displays actual compressive strength data for $n = 5$ specimens from each of the 23 initial batches of raw material purchased for an aircraft application. The first four lots of raw material were used in initial qualification. The mean and standard deviation of these twenty values are $\bar{x} = 143.95$ and $s = 8.29$, respectively. We assume that an acceptable probability of a Type 1 error is $\alpha = .001$, and obtain the values $k_1 = 3.7132$ and $k_2 = 1.4687$, from Tables 1 and 2, respectively. The acceptance limits are then calculated as

$$A_1 = \bar{x} - k_1 s = 113.2$$

and

$$A_2 = \bar{x} - k_2 s = 131.8.$$

Figure 4 show the data for all 23 lots, plotted against time. It can be seen that none of the means and minima of data from lots 5-23 fall below their respective acceptance limits; hence there is no indication provided to examine any of these lots further.

CONCLUSIONS

A saddlepoint approximation to the joint distribution of the mean and an extremum leads to a bivariate lot-acceptance chart which should prove useful in industrial applications where one would like protection against very low or high individual values, as well as against shifts in the mean. The approximation has been shown to be accurate for all sample sizes, and with the provided tables it is easy to use. A power analysis suggests that tests based on the mean and an extremum are typically at least comparable in power to tests based on the mean alone, and they can be substantially more powerful in detecting increases in standard deviation, and robust against spurious extreme values.

APPENDIX: DETAILS OF THE SADDLEPOINT APPROXIMATION

The approach which we will use begins with the observation that

$$\Pr\left(\bar{X} \leq t_2 | X_{(1)} = t_1\right) = \Pr\left[\frac{\sum_{i=1}^{n-1} U_i + t_1}{n} \leq t_2\right],$$

where the $\{U_i\}_{i=1}^{n-1}$ are *iid* truncated normal random variables, with support on the interval $[t_1, \infty)$. Let $U = \sum_{i=1}^{n-1} U_i$ denote this sum. The cumulant generating function of U_i, and its first two derivatives, are easily shown to be

$$\begin{aligned} K(t) &= \log[1 - \Phi(t_1 - t)] - \log[1 - \Phi(t_1)] + t^2/2 \\ K'(t) &= h(t_1 - t) + t, \text{ and} \\ K''(t) &= (t_1 - t)h(t_1 - t) - h^2(t_t - t) + 1. \end{aligned}$$

The saddlepoint approximation to the density of U is

$$f_U(s) \approx \frac{e^{-\lambda s + (n-1)K(\lambda)}}{\sqrt{2\pi(n-1)K''(\lambda)}}, \tag{9}$$

where λ is the (unique) root to the equation

$$K'(\lambda) = s/(n-1). \tag{10}$$

The approximate equality symbol ($\approx$) used above will appear in several places in this appendix. We take this opportunity to point out that the *only* approximation made in these derivations is the replacement of the true distribution of U with its saddlepoint approximation, as in (9).

The joint density of $\bar{X}$ and $X_{(1)}$ can be expressed as

$$\begin{aligned} f_{X_{(1)},\bar{X}}(t_1, t_2) &= \left[\frac{\mathrm{d}}{\mathrm{d}t_2} \Pr(\bar{X} \leq t_2 | X_{(1)} = t_1)\right] f_{X_{(1)}}(t_1) \\ &= n f_U(nt_2 - t_1) \left[n\phi(t_1)(1 - \Phi(t_1))^{n-1}\right] \end{aligned}$$

Substituting in (9), we have that

$$f_U(nt_2 - t_1) \approx \frac{e^{-\lambda(nt_2 - t_1)}}{\sqrt{2\pi(n-1)K''(\lambda)}} \frac{e^{\frac{n-1}{2}[\lambda^2 - (t_1 - \lambda)^2]}}{(2\pi)^{(n-1)/2} h^{n-1}(t_1 - \lambda)} (1 - \Phi(t_1))^{-(n-1)},$$

so

$$\begin{aligned} f_{X_{(1)},\bar{X}}(t_1, t_2) &\approx \frac{n^2 e^{-\lambda(nt_2 - t_1) + \frac{n-1}{2}[\lambda^2 - (t_1 - \lambda)^2] - \frac{t_1^2}{2}}}{(2\pi)^{(n+1)/2}\sqrt{(n-1)K''(\lambda)} h^{n-1}(t_1 - \lambda)} \\ &= \frac{n^2 \exp\left\{-\frac{n}{2}\left[t_1^2 - 2(t_1 - \lambda)(t_2 - t_1) + 2t_1(t_2 - t_1)\right]\right\}}{(2\pi)^{(n+1)/2}\sqrt{(n-1)K''(\lambda)} h^{n-1}(t_1 - \lambda)}. \end{aligned}$$

By subtracting t_1 from both sides of (10), one can see that the saddlepoint λ is the root of the equation

$$h(t_1 - \lambda) - (t_1 - \lambda) = \frac{n}{n-1}(t_2 - t_1). \tag{11}$$

Thus, $t_1 - \lambda$ is a function of $t_2 - t_1$, which we denote $q(t_2 - t_1)$. Since $K''(\lambda)$ is a function of $t_1 - \lambda$ alone, it can be regarded as a function of $t_2 - t_1$, say $L(t_2 - t_1)$. For the same reason, $h(t_1 - \lambda)$ is also a function of $t_2 - t_1$, which we will denote $\tilde{h}(t_2 - t_1)$.

This motivates the change of variables $Z = X_{(1)}$ and $W = \bar{X} - X_{(1)}$, which has unit Jacobian and which leads to

$$\begin{aligned} f_{W,Z}(w,z) &\approx \frac{n^2 \exp\left[-\frac{n}{2}\left(z^2 + 2zw - 2wq(w)\right)\right]}{(2\pi)^{(n+1)/2}\sqrt{L(w)}\tilde{h}^{n-1}(w)} \\ &= \frac{n^2 e^{-n(z+w)^2/2} e^{n(w^2+2wq(w))/2}}{(2\pi)^{(n+1)/2}\sqrt{L(w)(n-1)}\tilde{h}^{n-1}(w)}. \end{aligned} \tag{12}$$

The approximate density (12) need not integrate to one. We thus modify it by dividing by the normalizing constant

$$c_n = \int_0^\infty dw \int_{-\infty}^\infty dz\, f_{W,Z}(w,z) = \int_0^\infty dw \frac{n^{3/2}}{(2\pi)^{n/2}} \frac{e^{\frac{n}{2}\left[w^2+2wq(w)\right]}}{\sqrt{L(w)(n-1)}\tilde{h}^{n-1}(w)}. \tag{13}$$

By definition, $h(t) - t = wn/(n-1)$, and $h(t) - t$ is monotone. This suggests the change of variable

$$w = \frac{n-1}{n}(h(t) - t), \tag{14}$$

with Jacobian

$$dw = \frac{n-1}{n}\left|h^2(t) - th(t) - 1\right| dt. \tag{15}$$

Now we can write (13) as an integral in t, *eliminating the need to determine* λ:

$$\begin{aligned} c_n &= \frac{\sqrt{n-1}}{(2\pi)^{n/2}} \int_{-\infty}^\infty h^{-(n-1)}(t) \exp\left[\frac{(n-1)^2}{2n}(h(t) - t)^2 + (n-1)t(h(t) - t)\right] \\ &\quad \cdot\sqrt{1 - h^2(t) + th(t)}dt. \end{aligned} \tag{16}$$

Since

$$\begin{aligned} \Pr(X_{(1)} \le t_1 \text{ and } \bar{X} \le t_2) &= \Pr(Z \le t_1 \text{ and } W \le t_2 - t_1) \\ &\quad + \Pr(Z + W \le t_2 \text{ and } W \ge t_2 - t_1), \end{aligned}$$

we can express the approximate cumulative distribution of $X_{(1)}$ and $\bar{X}$ as the following sum of two integrals in the (W, Z) plane:

$$\begin{aligned} p &= \Pr(X_{(1)} \le t_1 \text{ and } \bar{X} \le t_2) \\ &\approx \left[\int_0^{t_2-t_1} dw \int_{-\infty}^{t_1} dz + \int_{t_2-t_1}^\infty dw \int_{-\infty}^{t_2-w} dz\right] \frac{1}{c_n} \frac{n^2 e^{-n(z+w)^2/2} e^{n(w^2+2wq(w))/2}}{(2\pi)^{(n+1)/2}\sqrt{(n-1)L(w)}\tilde{h}^{n-1}(w)}, \end{aligned} \tag{17}$$

where an obvious notation has been used to indicate the region of integration. As was the case in the derivation of c_n, integration in z can be performed explicitly, after which (17) becomes

$$\begin{aligned} p &= \frac{n^{3/2}}{c_n(2\pi)^{n/2}} \left\{\int_0^{t_2-t_1} \frac{e^{\frac{n}{2}\left[w^2+2wq(w)\right]}}{\sqrt{(n-1)L(w)}\tilde{h}^{n-1}(w)} \Phi\left[\sqrt{n}(t_1 + w)\right] dw \right. \\ &\quad \left. + \int_{t_2-t_1}^\infty \frac{e^{\frac{n}{2}\left[w^2+2wq(w)\right]}}{\sqrt{(n-1)L(w)}\tilde{h}^{n-1}(w)} \Phi(\sqrt{n}t_2) dw \right\}. \end{aligned} \tag{18}$$

Finally, we make the same change of variable (14,15) as in (16) to derive (1).

References

[1] Barndorff-Nielsen, O.E. and Cox, D.R. Asymptotic Techniques for Use in Statistics, New York: Chapman and Hall, 1989.

[2] Bissinger, B.H. and Mutry, V.N. "The Joint Probability Density Function of the Sample Mean and the Sample Maximum," Gujarat Statist. Rev., **8**, 7-14, 1981.

[3] Brickenkamp, C.E., Hasko, S., Natrella, M.G. Checking the Net Contents of Packaged Goods, NBS Handbook 133, Third Edition, Gaithersburg, MD: National Bureau of Standards, 1988.

[4] Croarkin, M.C. and Yang, G.L. "Acceptance Probabilities for a Sampling Procedure Based on the Mean and an Order Statistic," J. Res. Nat. Bur. Standards, **87**, 485-511, 1987.

[5] Stricklett, K.L. and Vangel, M.G. "Electric Motor Efficiency Testing Under the New Part 431 of Chapter II of Title 10, Code of Federal Regulations," NIST Technical Note TN 1422, Gaithersburg, MD: National Institute of Standards and Technology, 1997.

Table 1: Constants for an Acceptance Test Using the Mean and an Extremum (Values for Extremum; k_1)

α	Sample Size (n)								
	2	3	4	5	6	7	8	9	10
.5	0.7166	1.0254	1.2142	1.3498	1.4548	1.54	1.6113	1.6724	1.7258
.25	1.2887	1.5407	1.6972	1.8106	1.899	1.9711	2.0317	2.0838	2.1295
.10	1.8167	2.0249	2.1561	2.252	2.3272	2.3887	2.4407	2.4856	2.525
.05	2.1385	2.3239	2.442	2.5286	2.5967	2.6527	2.7	2.7411	2.7772
.025	2.4208	2.5888	2.6965	2.7758	2.8384	2.89	2.9337	2.9717	3.0052
.01	2.7526	2.9027	2.9997	3.0715	3.1283	3.1753	3.2153	3.25	3.2807
.005	2.9805	3.1198	3.2103	3.2775	3.3309	3.3751	3.4127	3.4455	3.4745
.0025	3.193	3.3232	3.4082	3.4716	3.522	3.5638	3.5995	3.6307	3.6582
.001	3.4549	3.5751	3.6541	3.7132	3.7603	3.7995	3.8331	3.8623	3.8883
.0005	3.6412	3.755	3.8301	3.8864	3.9314	3.969	4.0011	4.0292	4.0541
.00025	3.8188	3.927	3.9987	4.0526	4.0958	4.1319	4.1628	4.1898	4.2138
.0001	4.0421	4.1439	4.2117	4.2629	4.304	4.3384	4.3678	4.3936	4.4166
.00005	4.2035	4.3011	4.3664	4.4157	4.4554	4.4886	4.5172	4.5422	4.5644
.000025	4.3592	4.453	4.516	4.5637	4.6022	4.6344	4.662	4.6863	4.7079
.00001	4.5573	4.6466	4.7069	4.7527	4.7897	4.8206	4.8473	4.8707	4.8915

Table 2: Constants for an Acceptance Test Using the Mean and an Extremum (Values for Mean; k_2)

α	Sample Size (n)								
	2	3	4	5	6	7	8	9	10
.5	0.1472	0.1591	0.1539	0.1473	0.141	0.1354	0.1303	0.1258	0.1217
.25	0.6266	0.5421	0.4818	0.4382	0.4048	0.3782	0.3563	0.3379	0.3221
.10	1.0539	0.8836	0.7744	0.6978	0.6403	0.5951	0.5583	0.5276	0.5016
.05	1.3076	1.0868	0.9486	0.8525	0.7808	0.7246	0.679	0.6411	0.6089
.025	1.5266	1.2626	1.0995	0.9866	0.9026	0.8369	0.7838	0.7396	0.7022
.01	1.7804	1.4666	1.2747	1.1425	1.0443	0.9678	0.9059	0.8545	0.811
.005	1.9528	1.6054	1.3941	1.2488	1.1411	1.0571	0.9893	0.933	0.8854
.0025	2.1123	1.7341	1.5049	1.3475	1.2309	1.1401	1.0668	1.0061	0.9546
.001	2.3076	1.8919	1.6408	1.4687	1.3413	1.2422	1.1622	1.0959	1.0397
.0005	2.4457	2.0035	1.7371	1.5546	1.4196	1.3145	1.2298	1.1596	1.1002
.00025	2.5768	2.1097	1.8287	1.6363	1.4941	1.3835	1.2943	1.2203	1.1578
.0001	2.7411	2.2429	1.9436	1.739	1.5877	1.4701	1.3752	1.2966	1.2301
.00005	2.8595	2.3389	2.0266	1.813	1.6553	1.5326	1.4337	1.3517	1.2824
.000025	2.9734	2.4313	2.1065	1.8844	1.7204	1.5928	1.49	1.4048	1.3327
.00001	3.1179	2.5487	2.2079	1.9751	1.8031	1.6694	1.5616	1.4723	1.3968

Figure 1: $\tilde{F}_{X_{(1)},\bar{X}}(t_1, t_2)$, for $n = 5$. The broken line is the locus of (t_1, t_2) pairs for which $F_{X_{(1)}}(t_1) = F_{\bar{X}}(t_2)$.

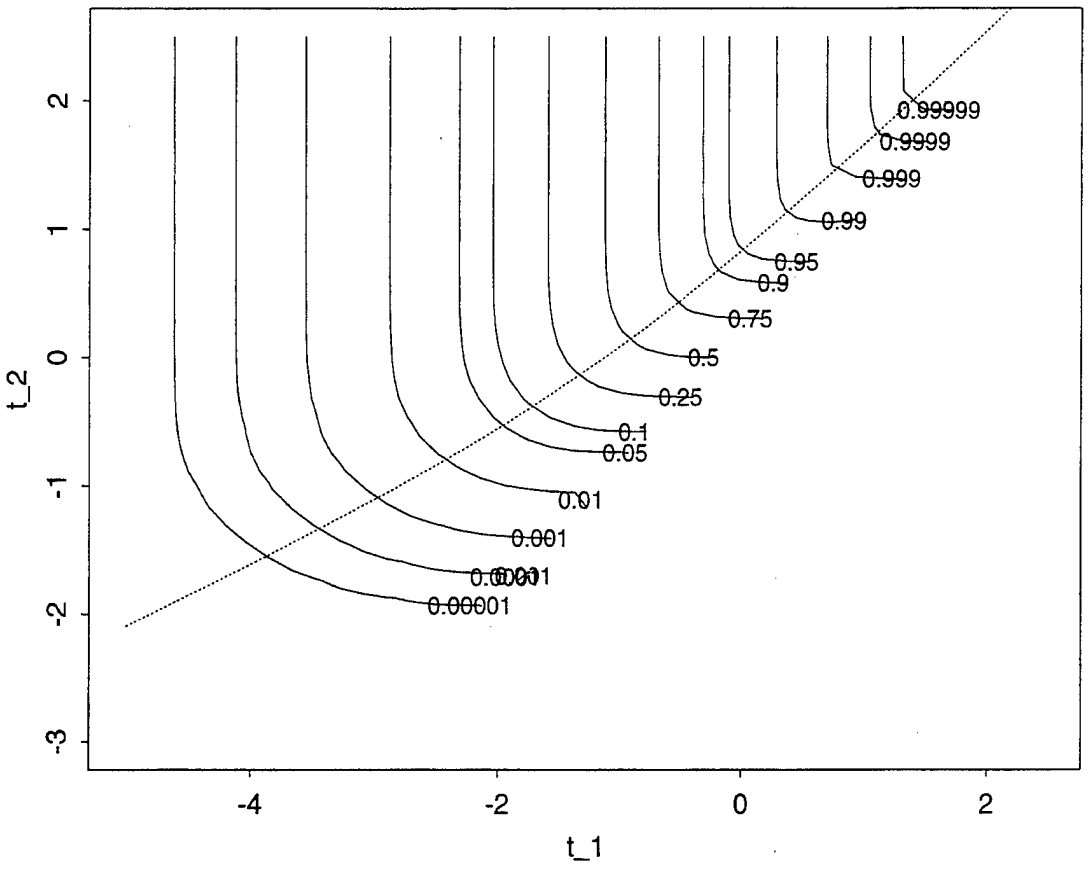


Figure 2: Solid lines are contours of $F_{X_{(1)},\bar{X}}(t_1, t_2)$, and broken lines are corresponding contours of $\tilde{F}_{X_{(1)},\bar{X}}(t_1, t_2)$, for $n = 2$.

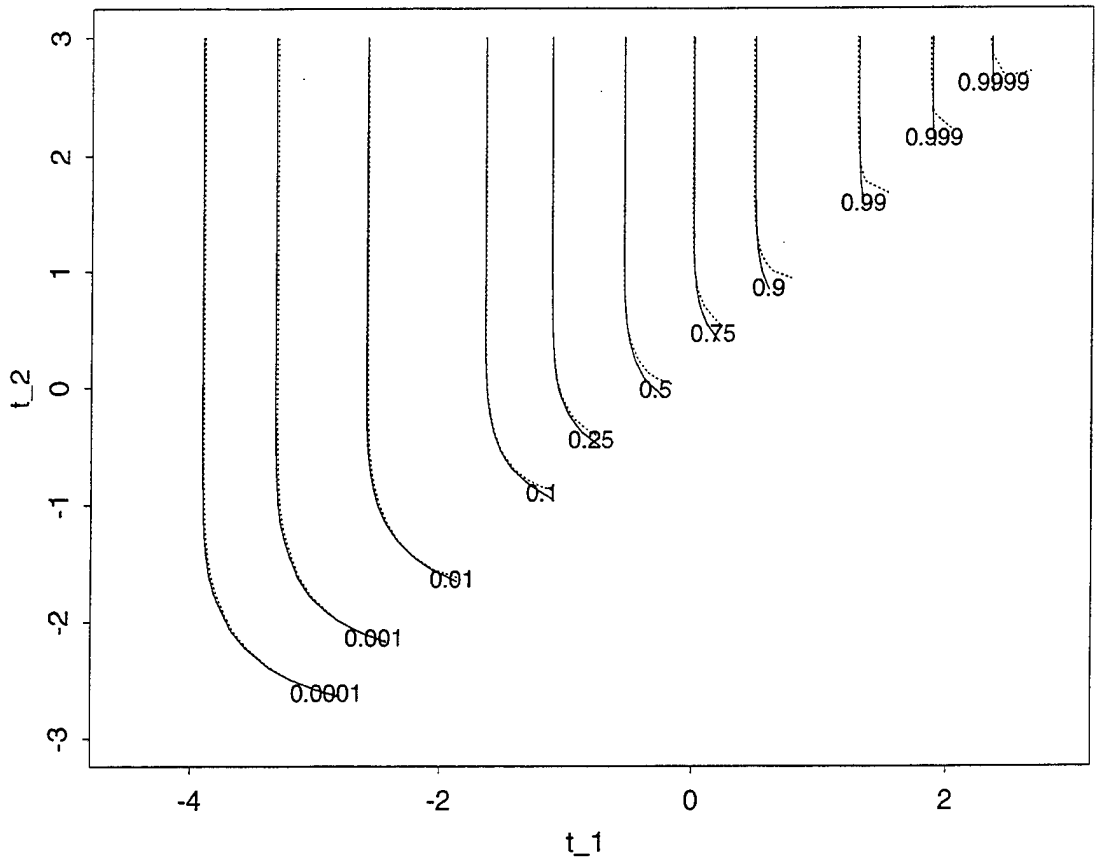


Figure 3: Difference in power $\pi_1(a,b) - \pi_2(a,b)$ between a test based on the mean and extremum, and a test based on the mean alone. The size is $\alpha = .001$, $n = 5$, and (3) holds.

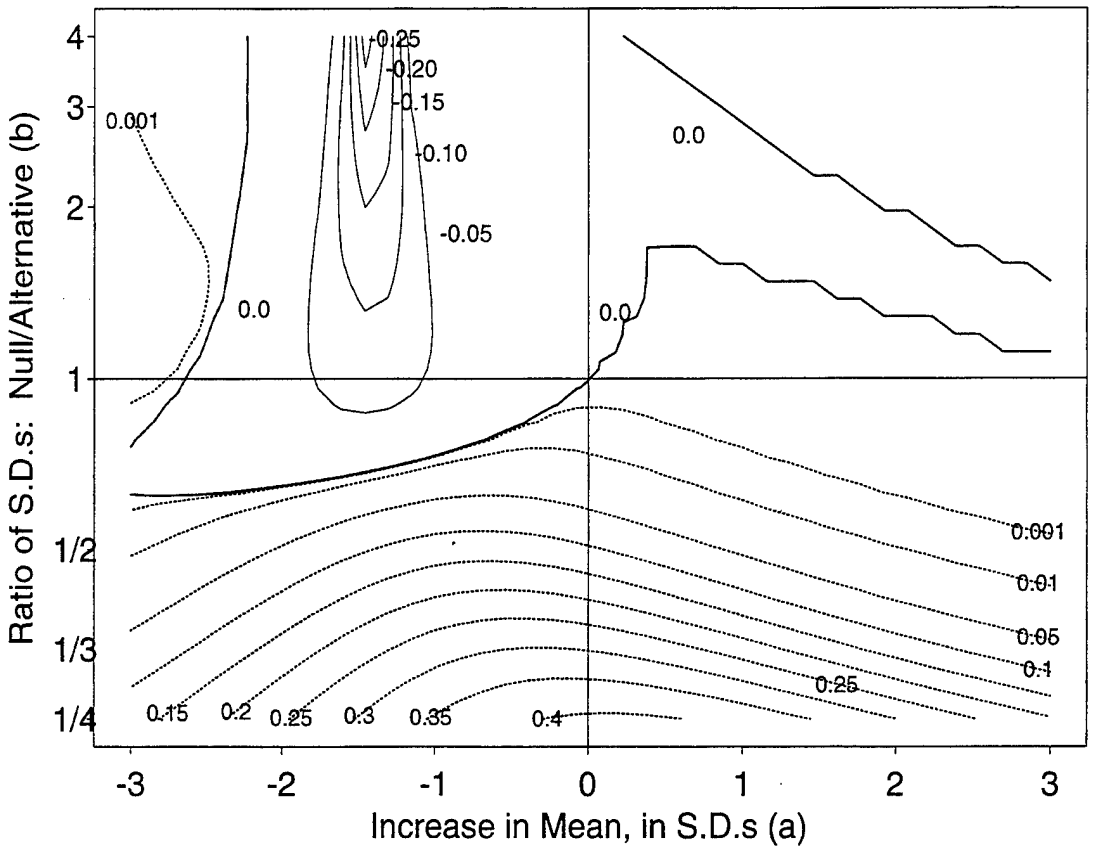
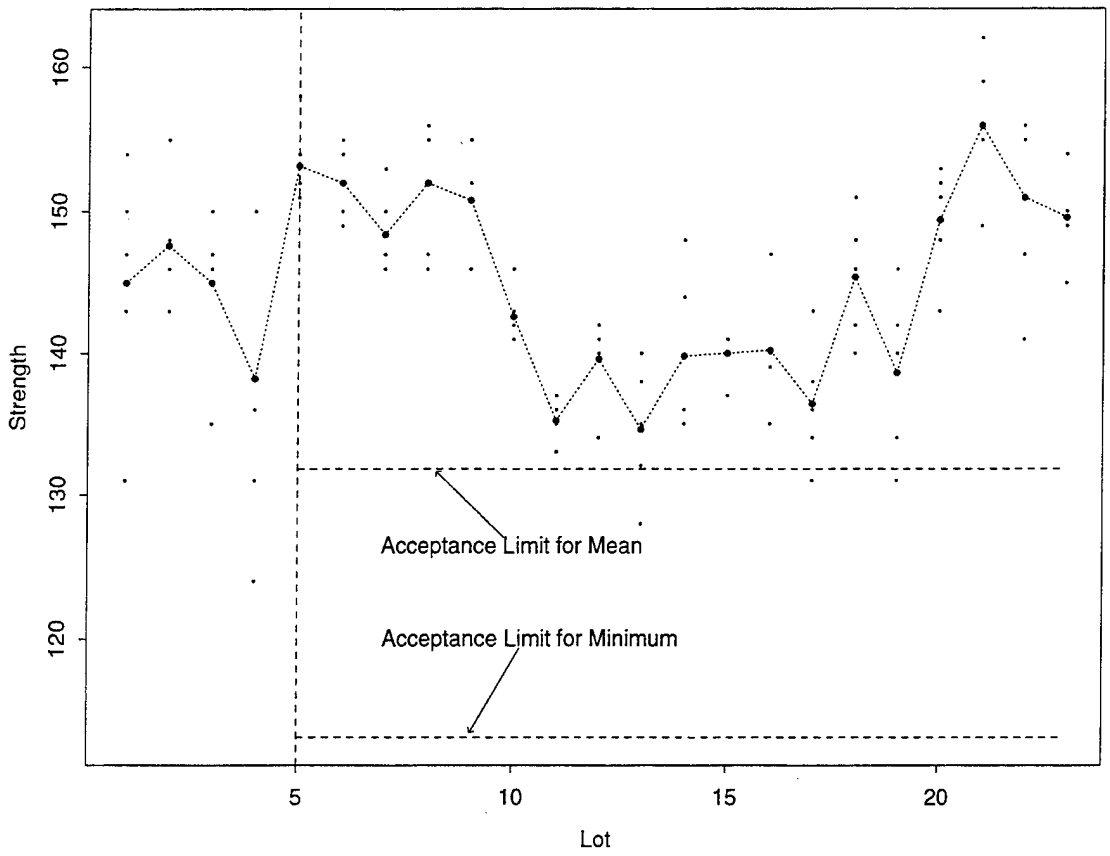


Figure 4: Example lot acceptance chart for compressive strength of a composite material ($n = 5, \alpha = .001$). The first four lots were used to set up the limits.



APPROXIMATE QUANTILES FOR THE MULTIVARIATE STUDENTIZED RANGE IN THE CASE OF THREE UNEQUAL GROUPS[1]

Otto Schwalb and James R. Thompson[2]
Department of Statistics
Rice University
6100 Main Street MS-138
Houston, TX 77005

ABSTRACT

In this paper, we develop equations which provide approximate quantiles for the multivariate studentized range in the case of three unequal groups. These equations are ready to use and are available via a website. In the equal groups case, our approximations are more accurate than existing approximations by a factor of ten. In the unequal groups case, there are no direct competitors to our approximations.

INTRODUCTION

The theory of multiple comparisons has a rich history in statistics. There is usually no debate about which multiple comparisons procedure is best for a given situation. In the multivariate case, for example, it is known that the T^2_{max} procedure (the simplest case of which is originally due to Roy (1953) and generalizations of which are attributed to Krishnaiah (1969)) will always outperform the equivariant procedures (among which are the procedures based upon Roy's root, the Pillai-Bartlett trace, the Hotelling-Lawley trace, and Wilks' Lambda) in cases where there are a fixed number of contrasts specified in advance and either confidence intervals or confidence regions (for particular parametric families) are desired. This claim is strong indeed and is perhaps not widely known, but it has been established for quite some time now. It requires more effort to demonstrate than just citing a few references here in the introduction (and some of the results required are fairly obscure), and so will be discussed below. That the T^2_{max} procedure possesses such outstanding properties is completely analogous to the relationship between Scheffe's procedure and the studentized range procedure in ANOVA. The reason that the T^2_{max} procedure has not received much attention seems to be due to the problems associated with obtaining its distribution in closed form. Siotani (1992) is one of the major investigators of the T^2_{max} procedure. His breakthrough in providing an equation for approximate quantiles for the general case of equal groups (after the closed form solution has remained an unsolved problem for nearly 40 years) is remarkable and should hopefully serve to bring renewed attention to a procedure which is so useful and easy to understand.

In this paper, we develop equations which provide approximate quantiles for the multivariate studentized range in the case of one-way MANOVA with three unequal groups. The equations apply to the following situations:

- quantiles: 0.8, 0.825, 0.85, ..., 0.925, 0.95, 0.975, 0.99 (e.g. for a significance level of $\alpha = 0.1$, we desire the $1 - \alpha = 0.9$ quantile)
- dimensions: 1, 2, 3, 4, 5
- group sizes: arbitrary group sizes between 4 and 100 (subject to the restriction that the largest group be no more than 4 times the size of the smallest group).

A JavaScript available at the web site:

[1]Approved for public release; distribution is unlimited.

[2]Otto Schwalb is a graduate student in the Statistics Department, Rice University, 6100 Main Street, MS-138, Houston, TX, 77005 (E-mail:schwalb@stat.rice.edu). James R. Thompson is Professor, Department of Statistics, Rice University, 6100 Main Street, MS-138, Houston, TX, 77005 (E-mail:thomp@stat.rice.edu).

allows the interested reader easy access to these quantiles. The importance of these approximate quantiles is that, until now, no quantiles have been available for the procedure in the case of unequal groups. Our approximations also apply to the case of equal groups, and, for the equal groups cases we consider, the accuracy of our approximations improves upon the accuracy reported by others by a factor of ten.

Section 2 treats the theory associated with the T^2_{max} procedure in the general case, as well as the special cases of two and three groups in 1-way MANOVA. Section 2 also demonstrates the claims made above concerning the performance of the T^2_{max} procedure with respect to the equivariant procedures. In section 3, the practical utility of the procedure is demonstrated with an example using the Fisher-Anderson Iris data (1939). The example illustrates how the T^2_{max} procedure is a strongly indicated choice in many realistic situations. Section 4 discusses the development of our equations. Most of the details are omitted from this section for the sake of brevity, but may be obtained from the first author. The details are straightforward, however.

THEORY

OVERVIEW

The derivation of the T^2_{max} procedure to follow is rather general, that is to say, very little of the theory to follow is specific to the 1-way case or the case of all pairwise comparisons. Parts of what follow can be found scattered throughout the literature, but the results are certainly not easy to find. For example, the results we synthesize on general confidence regions are very obscure. They are certainly beyond the reach of anyone who simply needs to use the procedures in an applied situation, but might have very little time to search the literature. Hopefully, the presentation here will make these results more accessible to a wider audience.

The MANOVA setup starts with a matrix Y which is $n \times p$ and whose rows are each distributed independently as p-variate normal vectors with common unknown covariance matrix Σ. It is further assumed that $E(Y) = X\beta$ where X is a design matrix of order $n \times l$ and β is the $l \times p$ matrix of unknown parameters. We assume that X has full column rank. We first define Roy's root and the related multivariate test statistics. We wish to test the hypothesis $H_0 : C\beta = 0$ where C is $r \times l$ and has full row rank. Define the hypothesis and error matrices (respectively) as

$$H = \hat{\beta}^T C^T W^{-1} C\hat{\beta} \quad \text{and} \quad E = Y^T \left(I - X(X^T X)^{-1} X^T\right) Y$$

where $\hat{\beta} = (X^T X)^{-1} X^T Y$ and $W = C(X^T X)^{-1} C^T$. Let the eigenvalues of HE^{-1} be denoted by $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$. Then Roy's root is λ_1, the Hotelling-Lawley trace is $\sum_{i=1}^p \lambda_i$, the Pillai-Bartlett trace is $\sum_{i=1}^p \lambda_i/(1+\lambda_i)$, and Wilks' lambda is $\prod_{i=1}^p 1/(1+\lambda_i)$. These statistics can be defined in similar ways using the eigenvalues of $H(E+H)^{-1}$, which tends to make the notation in the literature sometimes confusing. In fact, Roy's root seems to be defined in different ways depending on the author. We follow Wijsman (1979) and Kres (1983) in defining Roy's root as λ_1. For convenience, we will subsequently refer to these four statistics as the equivariant statistics.

For the T^2_{max} procedure, however, the hypothesis is stated in a different form as $H_0 : L\beta = 0$, where L is not required to have full row rank. The matrix $L\beta$ is an explicit representation of each and every contrast of interest for the problem at hand. Since anything of the form $a^T C\beta$, $\forall a \in \Re^r$, is protected by procedures based on the equivariant statistics, the family of linear combinations considered by those procedures is vast in comparison with the family of linear combinations considered by the T^2_{max} procedure. It should be apparent that the null hypothesis considered by the T^2_{max} procedure and the null hypothesis for the classical formulation are not equivalent. This might seem confusing, however, the problem of multiple comparisons deals with obtaining the best (usually in the sense of the smallest) set of confidence regions for a given parametric family. Within the classical framework then, we start with the hypothesis $H_0 : C\beta = 0$. Once a parametric family is specified (e.g. $a^T C\beta$, $\forall a \in \Re^r$), the discussion of competing multiple comparison procedures follows next. Such a discussion cannot be carried out prior

to specifying the parametric family of interest (see Wijsman (1979) for instance). With these subtleties in mind, the T^2_{max} procedure tests the null hypothesis that $H_0 : L\beta = 0$ vs $H_1 : L\beta \neq 0$ where L is a $t \times l$ matrix of contrasts. For example, in the case of 1-way MANOVA with 3 groups, $l = 3$ (the number of groups) and $t = 3$ (the number of pairwise comparisons). Applying a variation on the union-intersection approach (originally due to Roy (1953)), one sees that:

$$\{H_0 : L\beta = 0\} \Leftrightarrow \bigcap_{i=1}^{t} \{H_{0i} : L_{[i,\,]}\beta = 0\}$$

where $L_{[i,\,]}$ represents the ith row of the contrast matrix L (e.g. in the case of 1-way MANOVA with 3 groups under the "cell means" formulation of the design matrix X, $L_{[i,\,]} = (1,\ -1,\ 0)$ compares group 1 and group 2). Define $\delta_i^T := L_{[i,\,]}\beta$. As usual, the least squares estimators $\hat{\beta}$ are a linear combination of Y. By the Gauss-Markov Theorem, the minimum variance unbiased estimators for the δ_i's are $\hat{\delta}_i^T = L_{[i,\,]}\hat{\beta}$ which are also linear combinations of Y. It is easily seen that

$$\hat{\delta}_i \sim N_p(\delta_i, w_{ii}\Sigma)$$

where w_{ii} is the ith diagonal element of $L(X^TX)^{-1}L^T$. Let $\delta_i^* = \delta_i/\sqrt{w_{ii}}$ and $\hat{\delta}_i^* = \hat{\delta}_i/\sqrt{w_{ii}}$. Then

$$\hat{\delta}_i^* \sim N_p(\delta_i^*, \Sigma)$$

so that

$$T_i^2 = (\hat{\delta}_i^* - \delta_i^*)^T \left(\frac{E}{n-l}\right)^{-1} (\hat{\delta}_i^* - \delta_i^*)$$

is distributed as Hotelling's T^2 where E is the error matrix (defined above). The T^2_{max} statistic is then defined as

$$T^2_{max} = \max_{1 \leq i \leq t} T_i^2$$

The procedure rejects H_0 if any of the t Hotelling T^2's exceed $T^2_{max}(\alpha)$, where $T^2_{max}(\alpha)$ denotes the $1-\alpha$ quantile of the distribution of T^2_{max}, i.e. the value such that $P(T^2_{max} \leq T^2_{max}(\alpha)) = 1 - \alpha$. As will be discussed immediately, each T^2 which exceeds $T^2_{max}(\alpha)$ is flagged as a significant contrast. It should be noted here that we will speak of the comparison of T_i^2 with $T^2_{max}(\alpha)$, but our equations actually give values for $T_{max}(\alpha)$ due to the conventions which are already established for the univariate studentized range, $q(\alpha)$, and the equal-groups multivariate studentized range, $R_{MAX}(\alpha)$. Thus, in practice, one would need to take square roots of the T_i^2's before comparing them to $T_{max}(\alpha)$.

The T^2_{max} procedure has the intuitive appeal that it rejects iff $T_i^2 > T^2_{max}(\alpha)$ for some i, $i = 1, \ldots, t$, which occurs iff at least one of the p-dimensional confidence ellipsoids does not contain the 0-vector. That this follows is a simple consequence of the fact that $T^2_{max}(\alpha)$ is chosen to satisfy

$$P_{H_0}[\ T_i^2 \leq T^2_{max}(\alpha),\ \ i = 1, \ldots, t\ ] = 1 - \alpha$$

which is the same as writing

$$P_{H_0}[\ (\hat{\delta}_i^* - \delta_i^*)^T \left(\frac{E}{n-l}\right)^{-1} (\hat{\delta}_i^* - \delta_i^*) \leq T^2_{max}(\alpha),\ \ i = 1, \ldots, t\ ] = 1 - \alpha \quad .$$

Beyond simply rejecting, however, each T_i^2 which exceeds $T^2_{max}(\alpha)$ will be flagged as significant. For example, if $t = 3$ and we are given data and find that T_1^2 and T_2^2 exceed $T^2_{max}(\alpha)$ but T_3^2 does not, then we would flag δ_1^* and δ_2^* as the culprits. The practical interpretation of this is subject to the limitations of any confidence statement as follows. If repeated samples are taken and the 3 ellipsoids are constructed

each time, then $(1-\alpha) \times 100\%$ of the time *all* 3 of these ellipsoids will contain the true δ_i^*'s, $i = 1, 2, 3$. Equivalently, $(1-\alpha) \times 100\%$ of the time, all 3 of the T_i^2's will reflect the location of the true δ_i^*'s, $i = 1, 2, 3$, as indicated by their relationship to the $T^2_{max}(\alpha)$ quantile. For example, if δ_1^* and δ_2^* were truly non zero, but δ_3^* was zero, then, under repeated sampling, $(1-\alpha) \times 100\%$ of the time we would find that $T_1^2 > T^2_{max}(\alpha)$, $T_2^2 > T^2_{max}(\alpha)$, *and* $T_3^2 < T^2_{max}(\alpha)$.

In addition to the procedure described above, one can also construct confidence intervals for all pairwise contrasts. Specifically, it can be shown - see Krishnaiah (1969, p 126) - that confidence intervals for the T^2_{max} procedure are given by

$$c^T\delta_i \in c^T\hat{\delta}_i \pm \sqrt{T^2_{max}(\alpha)\, c^T E c\, w_{ii}/(n-l)} \quad , \quad \forall c \in \Re^p$$

Confidence intervals for the pairwise differences on each variable can be obtained by substituting $c_j^T = (0, \cdots, 0, 1, 0, \cdots, 0)$, where the 1 is in the jth position, for c in the previous equation. This leads to confidence intervals of the form

$$\delta_{ij} \in \hat{\delta}_{ij} \pm \sqrt{T^2_{max}(\alpha)\, e_{jj}\, w_{ii}/(n-l)} \quad , \quad \begin{matrix} i = 1, \ldots, t \\ j = 1, \ldots, p \end{matrix} \tag{1}$$

where δ_{ij} is the jth component of δ_i and e_{jj} is the jth diagonal element of E. Notice in constructing these intervals that whatever is to be gained over a 1-d battery of univariate procedures (the expression "1-d battery" will be described below in the example using the Iris data) depends completely upon the distribution of T^2_{max}. This follows since e_{jj} is just the univariate SSE on the jth dependent variable. The covariance structure within the data is largely reflected in the off-diagonal elements of E, and the off-diagonal elements completely dominate the tilting of the p-dimensional ellipsoids. But the off-diagonal elements are not used in (1), hence important information is being ignored if one looks at confidence intervals (along the dimensions of the variables as given) alone. This particular issue lies at the heart of the example we give involving the Iris data (below). We will revisit it once we get there.

CONFIDENCE INTERVALS

For the case of a fixed number of contrasts specified in advance, Krishnaiah (1969, p. 131) demonstrates that the T^2_{max} confidence intervals are smaller than those for Roy's root (analogous to the ellipsoid discussion below, the T^2_{max} confidence intervals provide protection for the parametric family of $a_i^T C \beta b$, for the pre-specified $\Re^r$ vectors $a_1, \ldots, a_t$ and $\forall b \in \Re^p$). In turn, Wijsman (1979) demonstrates that the confidence intervals based upon Roy's root are the smallest among the class of all equivariant procedures which provide protection for the parametric family $a^T C \beta b$, $\forall a \in \Re^r$ and $\forall b \in \Re^p$. This class of equivariant procedures includes those based on Wilks' Lambda, the Hotelling-Lawley trace, and the Pillai-Bartlett trace. Hence the confidence intervals based upon T^2_{max} are always shorter than those based upon these equivariant procedures when the number of contrasts considered is fixed and specified in advance.

CONFIDENCE REGIONS

We start this section by stating (with a reference) why the T^2_{max} ellipsoids are smaller than the ellipsoids based upon Roy's root. We also illustrate the relationship between Roy's root and T^2_{max} which dramatically emphasizes the conservative nature of Roy's procedure. After that, we cite the results which show that the Roy's root ellipsoids are the smallest (for all practical purposes) among the equivariant class of MANOVA procedures which provide protection for the parametric family $a^T C \beta$, $\forall a \in \Re^r$. The conclusion then is that the T^2_{max} ellipsoids will always be smaller than ellipsoids based on equivariant MANOVA procedures in all practical cases (provided we make additional mild restrictions to the equivariant class under consideration; one such sufficient condition is that the confidence regions in p space be convex).

The fact that the ellipsoids (for a fixed number of contrasts specified in advance) for the T^2_{max} procedure are always smaller than the ellipsoids based on Roy's root is an immediate consequence of the

fact that $T^2_{max}(\alpha)/(n-l) \leq \lambda_\alpha$ (see Krishnaiah (1969, p. 131)). In the presentation below, we illustrate the conservative nature of the equivariant procedures via a reexpression of the Roy's root ellipsoidal simultaneous confidence regions. For clarity here, we are using λ_α to denote the $1-\alpha$ quantile of Roy's root (i.e. $P(\lambda_1 \leq \lambda_\alpha) = 1-\alpha$ where we defined λ_1 at the beginning of the THEORY section); our λ_α coincides with the λ_α of Wijsman (1979) and Kres (1983).

The ellipsoids for Roy's root are given by (see Wijsman (1979), equation (2.4))

$$(a^TC\hat{\beta} - a^TC\beta)E^{-1}(a^TC\hat{\beta} - a^TC\beta)^T \leq \lambda_\alpha a^TC(X^TX)^{-1}C^Ta \quad , \quad \forall a \in \Re^r \tag{2}$$

where the C matrix is the full rank version of the contrast matrix for H_0 as described previously, and r is the rank of H_0 (for the 1-way case $r = l-1$). Now let $\lambda^*_\alpha = \lambda_\alpha(n-l)$. Rewrite (2) as

$$\frac{1}{a^TC(X^TX)^{-1}C^Ta}\,(a^TC\hat{\beta} - a^TC\beta)\left(\frac{E}{n-l}\right)^{-1}(a^TC\hat{\beta} - a^TC\beta)^T \leq \lambda^*_\alpha \quad , \quad \forall a \in \Re^r$$

which is equivalent to

$$\sup_{a\in\Re^r} T^2_a \leq \lambda^*_\alpha \quad ,$$

where

$$T^2_a := \frac{1}{a^TC(X^TX)^{-1}C^Ta}\,(a^TC\hat{\beta} - a^TC\beta)\left(\frac{E}{n-l}\right)^{-1}(a^TC\hat{\beta} - a^TC\beta)^T \quad .$$

We have shown in a different section that the T^2_{max} ellipsoids are given by

$$(\hat{\delta}^*_i - \delta^*_i)^T\left(\frac{E}{n-l}\right)^{-1}(\hat{\delta}^*_i - \delta^*_i) \leq T^2_{max}(\alpha) \quad , \quad i = 1,\ldots,t$$

which after some rearranging can be written

$$\frac{1}{a_i^TC(X^TX)^{-1}C^Ta_i}\,(a_i^TC\hat{\beta} - a_i^TC\beta)\left(\frac{E}{n-l}\right)^{-1}(a_i^TC\hat{\beta} - a_i^TC\beta)^T \leq T^2_{max}(\alpha) \quad , \quad i = 1,\ldots,t \tag{3}$$

where a_i^T is a row vector in $\Re^r$ chosen so that $a_i^TC = L_{[i,\,]}$ (it is always possible to choose such an a_i^T, since given L, one can choose C to be a matrix such that all the rows of L are in the span of C and C has full row rank). Rewrite (3) as

$$\max_{a\in\{a_1,a_2,\ldots,a_t\}} T^2_a \leq T^2_{max}(\alpha)$$

where

$$T^2_{a_i} := \frac{1}{a_i^TC(X^TX)^{-1}C^Ta_i}\,(a_i^TC\hat{\beta} - a_i^TC\beta)\left(\frac{E}{n-l}\right)^{-1}(a_i^TC\hat{\beta} - a_i^TC\beta)^T$$

And, expressed in this fashion, it becomes obvious that

$$T^2_{max} = \max_{a\in\{a_1,a_2,\ldots,a_t\}} T^2_a \leq \sup_{a\in\Re^r} T^2_a \quad . \tag{4}$$

The relationship as expressed in (4) clearly shows that the T^2_{max} procedure will always outperform Roy's procedure for the situation of a fixed number of contrasts specified in advance. It also emphasizes dramatically the conservative nature of the equivariant procedures, particularly when the number of contrasts t is very small.

Finally, Wijsman (1979) demonstrates that the simultaneous ellipsoidal confidence regions based upon Roy's root are the smallest among all equivariant procedures (which provide protection for the family $a^T C\beta, \forall\, a \in \Re^r$) when $p < l - r$, where l is the space spanned by the columns of X and r is the number of constraints in the null hypothesis. For the case of 1-way MANOVA with 3 groups, for example, $l = 3$ and $r = 2$, since the hypothesis testing in this case essentially asks whether the 3 p-variate location vectors for the groups can be adequately described by a single p-variate location vector plus noise (see the canonical form of the MANOVA model in Lehmann (1986)). Wijsman (1979) demonstrates further that the Roy's root ellipsoids are also the smallest in the equivariant class of procedures in the case where $p \geq l - r$, provided some very mild conditions are imposed upon the shapes of the confidence regions desired. Requiring the confidence regions to be convex is one example of such a mild condition. Another example is requiring the confidence regions to be connected and contain their "center" (i.e. contain the point estimate for the mean vector). From an applied point of view, it would be difficult to characterize these additional conditions as restrictions at all, and it would be fair to say that Wijsman's results bring a rich history of discussion on the subject to a close.

THE CASE OF TWO GROUPS

When there are two groups, there is only one contrast to consider. Let n_i be the sample size of group i and $n = n_1 + n_2$. Here $T^2_{max} = n_1 n_2/(n_1 + n_2)(\bar{x}_1 - \bar{x}_2)^T (n-2)E^{-1}(\bar{x}_1 - \bar{x}_2)$, which one may easily verify is the 2-sample Hotelling's T^2. Hence, $T^2_{max}(\alpha, n, p) = p(n-2)/(n-p-1)F_{p,n-p-1}(\alpha)$ and the quantiles are available in closed form. If $T^2_{max} > T^2_{max}(\alpha)$, then the procedure rejects.

THE CASE OF THREE GROUPS

In the case of three groups, there are 3 contrasts, and the distribution is no longer available in closed form although there has been a great deal of effort in that direction (see Siotani (1959) and (1992)). Using the "cell means" formulation of the design one is led to $(X^T X)^{-1} = diag(1/n_1, 1/n_2, 1/n_3)$ and the contrast matrix

$$L = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}$$

which leads to

$$L(X^T X)^{-1} L^T = \begin{pmatrix} 1/n_1 + 1/n_2 & 1/n_1 & -1/n_2 \\ 1/n_1 & 1/n_1 + 1/n_3 & 1/n_3 \\ -1/n_2 & 1/n_3 & 1/n_2 + 1/n_3 \end{pmatrix}$$

So in this simple case, $w_{11} = 1/n_1 + 1/n_2$, $w_{22} = 1/n_1 + 1/n_3$, and $w_{33} = 1/n_2 + 1/n_3$. Also, $\delta_1 = \mu_1 - \mu_2$, $\delta_2 = \mu_1 - \mu_3$, and $\delta_3 = \mu_2 - \mu_3$. It is also clear (this is similar to the general case already discussed) that $\hat{\delta}_1 = (\bar{x}_1 - \bar{x}_2) \sim N_p(\delta_1, w_{11}\Sigma)$. Rewrite this as $\hat{\delta}_1^* \sim N_p(\delta_1^*, \Sigma)$ (where $\hat{\delta}_1^* := \hat{\delta}_1/\sqrt{w_{11}}$). This yields

$$T_1^2 = (\hat{\delta}_1^* - \delta_1^*)^T \left(\frac{E}{n-l}\right)^{-1} (\hat{\delta}_1^* - \delta_1^*) = \frac{1}{w_{11}} (\bar{x}_1 - \bar{x}_2)^T \left(\frac{E}{n-l}\right)^{-1} (\bar{x}_1 - \bar{x}_2)$$

Similarly for T_2^2 and T_3^2. Then T^2_{max} is just the largest of these 3 statistics. The procedure rejects, then, if any of the three T_i^2's exceed $T^2_{max}(\alpha, n_1, n_2, n_3, p)$. Additionally (as explained above) each T_i^2 which exceeds the $T^2_{max}(\alpha)$ quantile is flagged as significant. As a point of reference it is helpful to note that if the group sizes are equal, then

$$T_{max}(\alpha, n_1, n_1, n_1, 1) = \frac{1}{\sqrt{2}}\, q(\alpha, 3, n-3)$$

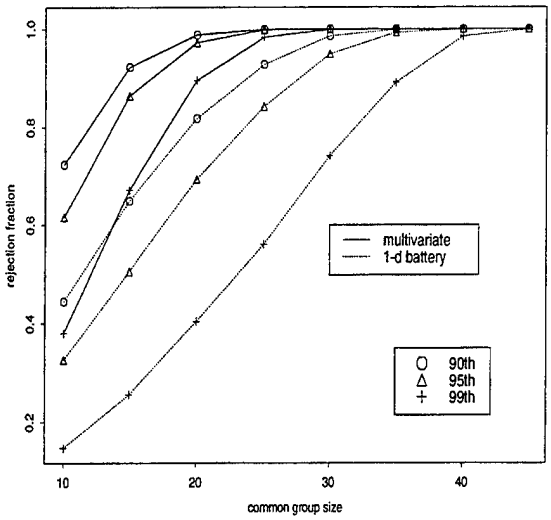


Figure 1: Plotted are the fractions of the time (out of 10,000 resamples) when the given procedure found the (shifted) versicolor-virginica difference.

where $q(\alpha, 3, n-3)$ is the $1-\alpha$ quantile of the univariate studentized range with 3 groups and $n-3$ error degrees of freedom ($n := n_1 + n_1 + n_1$). And also for the case of equal group sizes, it is helpful to note that

$$T_{max}(\alpha, n_1, n_1, n_1, p) = \frac{1}{\sqrt{2}} R_{MAX}(\alpha, m, \nu, p)$$

where R_{MAX} is the *multivariate studentized range* as defined in Siotani (1992) with parameters m and ν. Under Siotani's definition $\nu = \sum_i n_i - 3$ and $m = 3$.

PRACTICAL UTILITY OF THE T^2_{max} PROCEDURE

It is a characteristic of the MANOVA problem that no single procedure is going to be the best for all situations. With this in mind, it seems appropriate to demonstrate the practical utility of the T^2_{max} procedure by comparing its performance with a reasonable competitor. The fact that the T^2_{max} procedure will always outperform Roy's root and, in turn, the other equivariant test statistics (for the all pairwise comparisons situation) clearly rules out procedures based on those statistics as competition. It is clear that a battery of univariate procedures protected against type I inflation via the Bonferroni method is a reasonable competitor. The studentized range procedure is known to provide the shortest confidence intervals (of the available procedures) for all pairwise comparisons for equal groups in the univariate case, and it is a remarkable fact that the procedure is conservative when the groups are not equal. Thus a battery (across dimensions) of studentized range procedures will be used. It should be noted in passing that there are situations where the studentized range 1-d battery will find differences when the T^2_{max} procedure does not, and vice versa. It should also be pointed out here that for all the cases covered by our approximations for T^2_{max}, confidence intervals based on the studentized range 1-d battery will always be shorter than confidence intervals based on T^2_{max}. Nonetheless, by comparing each T^2_i with $T^2_{max}(\alpha)$ there are many situations where the T^2_{max} procedure will find group differences when the Bonferroni-studentized range procedure will not, which this example will show.

We consider the Fisher-Anderson Iris data (1939). This data set consists of 3 groups with 4 dimensions. There is no problem in declaring the setosa group to be different from the versicolor and virginica groups. It may be surprising that the 1-d studentized range battery also finds the versicolor-virginica difference

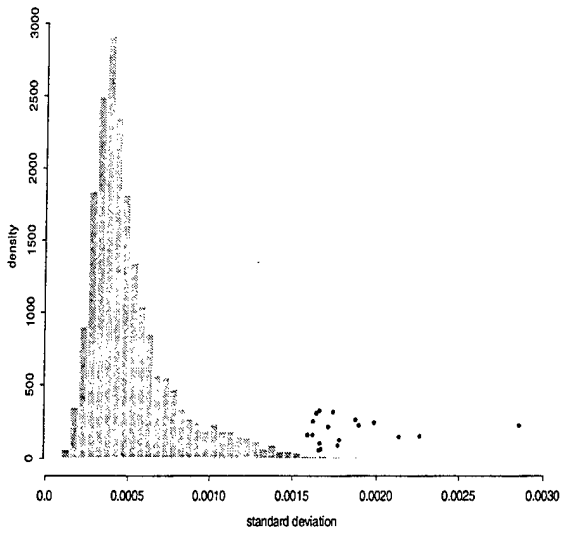


Figure 2: Histogram of all 4500 standard deviations for the validation runs. Overlaid on the plot are the largest 18 standard deviations.

quite easily, even at very high levels of significance. However, marginal plots of petal length and petal width clearly show these two groups to be different. Shifting the versicolor group closer to the virginica group provides a more interesting example for the procedures at hand. We shift the versicolor group by adding $(1, 0, 1.5, 0.5)$ to each observation where the coordinates are (sepal l., sepal w., petal l., petal w.). It might be even more surprising that the 1-d battery still finds the versicolor-virginica difference to be highly significant.

Now suppose we took random samples without replacement of size $n_0 < 50$ (say $n_0 = 45$, for now) from each group (we stick with equal samples since this is where the univariate studentized range procedure is the most powerful) and each time applied (1) the 1-d battery and (2) the T^2_{max} procedure. We can record whether each procedure finds the versicolor-virginica difference for each sample. We do this some large number of times. We can then repeat the whole setup for $n_0 = 40, 35, 30, \ldots$, etc. to get some idea of how the procedures perform as we decrease the sample size.

We carry this out $10,000$ times at each value of n_0 and summarize the results in Figure 1. Even with only a subsample of 25 points from each group, the T^2_{max} procedure suffers practically no loss in performance, declaring the versicolor-virginica difference about 98% of the time using the $\alpha = 0.01$ critical value. On the other hand, the 1-d battery declares the versicolor-virginica difference only about 56% of the time using its (conservative) $\alpha = 0.01$ critical value for subsamples of size 25 from each group. A rotating 3-d plot of the shifted data in software such as Xgobi illustrates dramatically how it is the covariance structure which separates the groups. The battery of 1-d tests is essentially blind to this type of multivariate structure. In all fairness, the T^2_{max} procedure would be blind in the case where the difference in the groups is small along a single dimension and many redundant dimensions (which contribute no help through their covariance structure - independence as one example) with no differences are included. Nevertheless, this example using the Iris data demonstrates the practical utility of the T^2_{max} procedure.

Returning to an issue mentioned earlier, what lies at the heart of this example is the fact that

$$\frac{1}{\sqrt{2}} q(\alpha/4, 3, n-3) < T_{max}(\alpha, n_1, n_1, n_1, 4)$$

so that the intervals in (1) are always smaller for the 1-d battery of studentized range procedures than they are for the T^2_{max} procedure. Hence, the only way to outperform the 1-d battery is by taking advantage

of the multivariate structure as we have done here (i.e. in comparing the T_i^2's to $T_{max}^2(\alpha)$ we are using the multivariate information which is contained in the off-diagonals of E). As mentioned above, for all the cases our equations cover,

$$\frac{1}{\sqrt{2}}q(\alpha/p, 3, n-3) < T_{max}(\alpha, n_1, n_1, n_1, p)$$

so that the 1-d studentized range battery always leads to shorter confidence intervals than T_{max}^2. This leads us to conjecture that in the 1-way case, the 1-d studentized range battery will always lead to shorter confidence intervals than T_{max}^2. This conjecture is clearly of great relevance from a practical standpoint.

DEVELOPING THE EQUATIONS

The equations we developed are based upon Monte Carlo simulation of the quantiles of T_{max}^2. It would have been possible of course to provide the reader with FORTRAN code which performs simulations of T_{max}^2 in a very general setting (e.g. for an arbitrary design X and an arbitrary contrast matrix L). However, it was decided that such a code would not be so widely used due to the fact that it requires the user to wait for a simulated quantile (the waiting time could be as much as an hour or so). To overcome this liability, it was decided to provide approximate quantiles in some format which does not require the user to wait. One immediately thinks of tables (of approximate quantiles) as a potential such format. One prohibitive difficulty of providing tables in a very general setting is that the completely general T_{max}^2 formulation is almost as general as the (multivariate) linear model. To make any headway at all in a table-type direction, one must consider particular cases of the general formulation of T_{max}^2.

The simplest case of T_{max}^2 is the case of one-way MANOVA. For the unequal groups case, however, one-way MANOVA is still too general, and one must restrict the generality even further to make any progress. Within one-way MANOVA, the natural place to start is 3 unequal groups, since quantiles for 2 unequal groups are available in closed form (this was described above). For 3 groups of unequal sizes, there are really too many combinations of the group sizes to be able to construct a table (of approximate quantiles) within any reasonable number of pages. Based on this difficulty, it was decided that we should offer the user a set of equations rather than a lengthy table. Not only can the equations provide a more concise summary of the approximate quantiles, but they can also interpolate (approximate) quantiles at which no simulations have been performed. Further, the user is not required to wait at all. Thus, equations for approximate quantiles of T_{max}^2 seemed the appropriate format; this is the format we provide for the case of 3 unequal groups.

The equations were developed as follows: (a) we chose a (large) grid, say $\mathcal{G}$, of (n_1, n_2, n_3) triples at which to perform simulations of T_{max}^2 and (b) after simulating quantiles at each point of $\mathcal{G}$, we fit (interpolating) equations to these simulated quantiles. We will refer to the simulations used in constructing these equations as the *model-building* simulations. The fitting of these equations is described in the first subsection (ABRIDGED SUMMARY OF THE FITTING PROCESS) below.

To obtain some idea of the accuracy of our equations, we performed additional simulations. These additional simulations will be referred to as the *validation* runs. In the validation runs, we performed a much larger number of simulations over a much smaller subset of points of $\mathcal{G}$. This is described in the second subsection (ASSESSING THE ACCURACY OF THE APPROXIMATIONS) below.

ABRIDGED SUMMARY OF THE FITTING PROCESS

It was necessary to restrict the approximations to the region (a) $4 \leq n_i \leq 100$, $i = 1, 2, 3$, and (b) $\max_{i\in\{1,2,3\}}(n_i) \leq 4\min_{i\in\{1,2,3\}}(n_i)$ in order to reduce the number of (n_1, n_2, n_3) triples considered to something manageable. Even with this reduction, 9357 (n_1, n_2, n_3) triples (about 5,000 distinct points) were used in the model-building simulations.

For a given dimension (recall that dimension corresponds to p), at least 100,000 simulations per point (or triple) were used, more simulations being required for the higher dimensions . Using an adaptive number of simulations was attempted but proved to be too slow and unstable. Thus the number of simulations per point had to be fixed for each dimension.

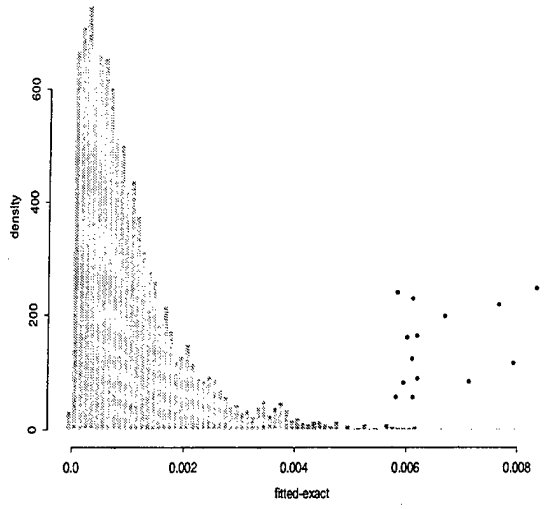


Figure 3: Histogram of the (absolute values) of all 4500 of the fitted values - the "exact" values. Overlaid on the plot are the 14 largest absolute errors.

It became clear from the inspection of many scatter plots that the function might be well approximated by a linear combination of terms like $1/n_1$, $1/n_2$, etc. Least squares was used to fit equations for each quantile and each value of p. Thus 45 equations were fit (5 dimensions: $p = 1, \ldots, 5$, and 9 quantiles: $0.8, 0.825, 0.85, \ldots, 0.925, 0.95, 0.975, 0.99$). The equations took the form

$$\sqrt{T^2_{max}(n_1, n_2, n_3)} = a_0 + \sum_{i=1}^{3}\sum_{s=1}^{k} a_{is}\left(\frac{1}{n_i}\right)^s + \sum_{1\le i<j\le 3}\sum_{s=1}^{k} a_{ijs}\left(\frac{1}{n_i n_j}\right)^s + \sum_{s=1}^{k} a_s\left(\frac{1}{n_1 n_2 n_3}\right)^s \qquad (5)$$

where k denotes the order of the polynomial used (we eventually arrived at $k = 4$). This form was determined via a standard regression approach. A second order model was actually enough to remove all lack-of-fit; however, adding the third and fourth order terms provided enormous improvements in accuracy (accuracy is discussed below). It is well known that a Laurent series expansion can sometimes lead to a more parsimonious expression with fewer coefficients. While a second order Laurent series expansion provided similar regression results, the accuracy of those equations proved to be much worse than the fourth order model (5), not to mention the fact that the Laurent series expansion required just as many coefficients.

ASSESSING THE ACCURACY OF THE APPROXIMATIONS

To evaluate the accuracy of the fitted equations, 100 of the original 9357 points were chosen at random (subject to some restrictions requiring that more points be chosen from the region of the smaller sample sizes, due to the fact that the region of the smaller sample sizes is harder to fit). At each of these 100 points, at least $1,000,000$ simulations were performed (again, more simulations were performed for higher dimensions). Each point was replicated 10 times in order to provide an estimate of variability at each of the 100 points. As an example, one of the 100 random points chosen was $(n_1, n_2, n_3) = (8, 12, 29)$. For $p = 1$, $1,000,000$ simulations of T^2_{max} were performed with $(n_1, n_2, n_3) = (8, 12, 29)$. The 9 sample quantiles $0.8, 0.825, \ldots, 0.95, 0.975, 0.99$ were recorded. This gave estimates of the true quantiles, call these estimates $T^2_{max}(\alpha, 8, 12, 29, 1)_{(1)}$, $\alpha = 0.2, 0.175, \ldots, 0.025, 0.01$. We repeat this 9 more times to get $T^2_{max}(\alpha, 8, 12, 29, 1)_{(2)}, \ldots, T^2_{max}(\alpha, 8, 12, 29, 1)_{(10)}$, $\alpha = 0.2, 0.175, \ldots, 0.025, 0.01$. Then (use $\alpha = 0.2$ for

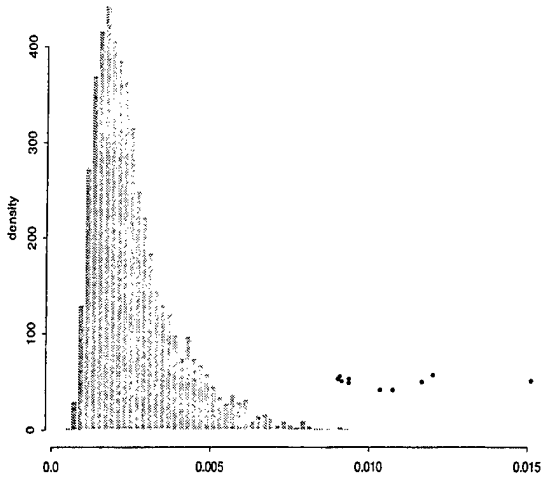


Figure 4: Histogram of 3*(standard deviation of "exact") + |fitted-"exact"|. Overlaid on the plot are the 10 largest such values.

example) we use the average of $T^2_{max}(0.2, 8, 12, 29, 1)_{(1)}, \ldots, T^2_{max}(0.2, 8, 12, 29, 1)_{(10)}$ as our estimate of the true quantile (we will refer to this average as the "exact" quantile). We use the standard deviation of $T^2_{max}(0.2, 8, 12, 29, 1)_{(1)}, \ldots, T^2_{max}(0.2, 8, 12, 29, 1)_{(10)}$ divided by $\sqrt{10}$ as an estimate of the standard deviation of the "exact" quantiles.

The increase in the number of simulations per point with the increase in dimension was adequate to keep these standard deviations constant across dimensions. And, again for brevity, a histogram of the standard deviations of the "exact" quantiles for all 4500 validation points (i.e. 100 points $\times$ 9 quantiles $\times$ 5 dimensions) is given in Figure 2. The largest standard deviation is 0.002853, but it is clear from the histogram that performance is generally much better than that. These "exact" estimates will be referred to as the *validation runs.*

To estimate the accuracy of the fitted equations, we calculate fitted-"exact" for each of the 100 validation points. A histogram of the absolute values of these results is given in Figure 3. The largest absolute error (i.e. |fitted - exact|) is 0.0083. The absolute errors rarely exceed 0.006, which can also be observed from the figure. In fact, the absolute errors never exceed 0.006 for $p = 1$ and $p = 2$. To be very conservative, if we believed that the true quantile were as far as ± 3 standard deviations from the "exact" quantile, then the furthest our fitted values could be from the truth would be *3*(standard deviation of "exact") + | fitted-"exact" |*. A histogram of these values is given in figure 4. The largest such value is 0.0151, but the performance is clearly better than that in general.

We may use the same analysis as in the preceding paragraph on the results Siotani (1992) reports. Due to his definitions, we must rescale what he reports by $\sqrt{2}$ (the rescaling is clearly indicated in the table below). Nevertheless, a direct comparison between his results and our results is appropriate. The worst case he does report is for $p = 5$ and $\alpha = 0.01$; we reproduce this case here (the first four columns here are taken from Siotani's Table 1 and Table 3).

ν	fitted	"exact"	s.d.("exact")	fitted-"exact"	$3 * \frac{s.d.}{\sqrt{2}} + \left\lvert \frac{f-e}{\sqrt{2}} \right\rvert$
10	14.359	14.31	0.297	0.049	0.6647
20	8.340	8.339	0.077	0.001	0.1641
60	6.551	6.548	0.051	0.003	0.1103

The 0.6647 value in the last column of this table compares directly to our "worst case" value of 0.0151.

Giving Siotani more than the benefit of the doubt, 0.1641 is his second worst value which is still more than 10 times larger than our worst value of 0.0151.

COMPUTING PLATFORM AND OTHER SOFTWARE

The far majority of the computing for these simulations was done on 2 computers, named *student* and *gauss*. *Student* is a Silicon Graphics Indigo 2 High Impact Workstation, and *Gauss* is a Pentium Pro 150 running FreeBSD 2.2. The running times were roughly 3 weeks for the validation runs and roughly 2.5 weeks for the model building runs. The computer code for the simulations consisted of roughly 300 lines of FORTRAN. The code calls a solve routine from LAPACK, as well as a few basic routines from the BLAS. The 45 separate regressions were carried out using ordinary proc reg in SAS, while diagnostic plots for the regressions (as well all of the plots in this paper) were done in Splus. The code used for the *random number generators* was written by Barry Brown, James Lovato, and Kathy Russell. It is all contained in the "randlib.f" library. The latest version (version 1.3) of the library is now available at *http://odin.mdacc.tmc.edu/anonftp/source.html.*

ACKNOWLEDGEMENTS

The first half of this work was supported by NIH Grant 5 R01 GM53545-02. The second half was supported by U.S. Army Research Office DAAH04-95-1 0665. The authors would like to thank Marek Kimmel, Dennis Cox, David Andrews, David Scott and Kathy Ensor for their help and support.

REFERENCES

——Anderson, E. (1939). "The Irises of the Gaspé Peninsula," *Bulletin of the American Iris Society*, **59**, 2-5.

——Kres, Heinz (1983). *Statistical Tables for Multivariate Analysis*, Springer-Verlag.

——Krishnaiah, P.R. (1969). "Simultaneous Test Procedures under General MANOVA Models." *Multivariate Analysis II*, 121-143.

——Lehmann, E.L. (1986). *Testing Statistical Hypotheses*, Wadsworth & Brooks/Cole.

——Roy, S.N. and Bose, R.C. (1953). "Simultaneous Confidence Interval Estimation." *Annals of Mathematical Statistics*, **24**, 513-536.

——Siotani, M. (1959)."The extreme value of the generalized distances of the individual points in the multivariate normal sample." *Annals of the Institute of Statistical Mathematics*, **76**, 725-728.

——Siotani, M. and Seo, Takashi (1992). "The multivariate studentized range and its upper percentiles." *Journal of the Japan Statistical Society*, **22**, 123-137.

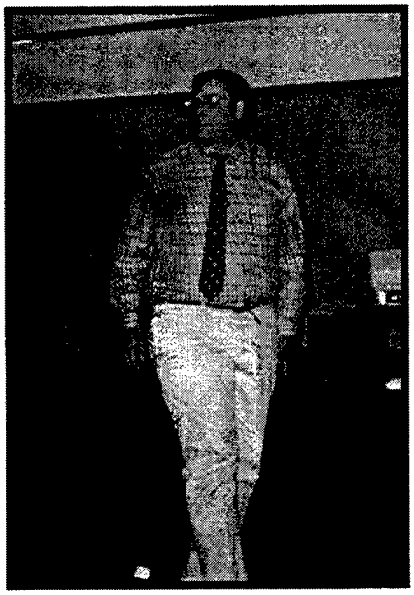
——Wijsman, Robert A. (1979). "Constructing all simultaneous confidence sets in a given class, with applications to MANOVA", *Annals of Statistics*, **7**, 1003-1018.

APPENDIX:

CONFERENCE SNAPSHOTS

INTENTIONALLY LEFT BLANK.

Left: George W. Johnson Center



Right: Wegman, conference host

Far Left: Conover, 1997 Army Wilks Award Winner; Standing: Thomas Hettmansperger, keynote speaker

Left to right, Army Wilks Award winners in attendance: Tang, Harris, Conover, Thompson, Sethuraman

INTENTIONALLY LEFT BLANK.

ATTENDANCE LIST

U.S. ARMY CONFERENCE ON APPLIED STATISTICS
22–24 OCTOBER 1997

Henry C. Alberts
5842 Hilldon St.
McLean, VA 22101

Kragg P. Kysor
P.O. Box 685
Edgewood, MD 21040

Edward J. Wegman
10821 Burr Oak Way
Burke, VA 22015

Patricia K. Joyce
Center for Comp. Stat.
George Mason Univ.
Fairfax, VA 22030

Donald A. Berry
ISDS Box 90251
Duke University
Durham, NC 27708

Barry Bodt
U.S. Army Research Laboratory
ATTN: AMSRL-IS-CI
Aberdeen Proving Ground, MD
21005-5067

Ann E. M. Brodeen
U.S. Army Research Laboratory
ATTN: AMSRL-IS-CI
Aberdeen Proving Ground, MD
21005-5067

Barbara D. Broome
U.S. Army Research Laboratory
ATTN: AMSRL-IS-CI (Bldg. 116A)
Aberdeen Proving Ground, MD
21005-5067

Aivars K. Celmiņš
U.S. Army Research Laboratory
ATTN: AMSRL IS-CI (Bldg. 116A)
Aberdeen Proving Ground, MD
21005-5067

Dr. Jeffrey B. Birch
Dept. of Stats. 210c Hutcheson Hall
VA Tech
Blacksburg, VA 24061

Dr. Barney Bissinger
281 W. Main St.
Middletown, PA 17057

Robert J. Burge
Div. of Biometrics, Bldg. 83
Walter Reed Army Inst. of Research
Washington, DC 20307

Jagdish Chandra
U.S. Army Research Laboratory
Aberdeen Proving Ground, MD
21005-5066

W. J. Conover
College of Bus. Admin.
Texas Tech. Univ.
Lubbock, TX 79409

D. H. Frank
336 Stright Hall
Indiana Univ. of PA
Indiana, PA 15705

Frank B. Gray
HQ AFOTEC/CN
8500 Gibson Blvd. SE
Kirtland AFB, NM 87117

Jim X. Chen
4212 Sideburn Rd.
Fairfax, VA 22030

Kyong Park
ERDGC, SCBRD-RTM
Bldg. E3330, Rm. 120
APG-EA, MD 21010

John Lyons
Johns Hopkins Laboratory/APL
Bldg. 24, Rm. E278
Laurel, MD 20723

Paul Pedersen
Los Alamos National Laboratory
MS B265
Los Alamos, NM 87545

Frances J. Morawski
Defense Information Systems Agency/D74
5600 Columbia Pike, 5th Floor
Falls Church, VA 22041

Adi Wilhelm
Center for Computational Statistics
George Mason Univ.
Fairfax, VA 22030

Cindy Krieger
107 Blueberry Ct.
Stafford, VA 22554

James R. Thompson
Dept. of Statistics, 6100 S. Main
Rice University
Houston, TX 77005

David W. Webb
U.S. Army Research Laboratory
ATTN: AMSRL-WM-PB
Aberdeen Proving Ground, MD
21005-5069

David Cruess
6123 Camelback Lane
Columbia, MD 21045

Kwang-Su Yang
Statistics-CSI
George Mason University
Fairfax, VA 22030

Thomas Sudkamp
Dept. of Computer Science
Wright State University
Dayton, OH 45435

Edward J. Mulrow
1061 N. George Mason Dr.
Arlington, VA 22205

Juergen Symanzik
Center for Computational Statistics
George Mason University
Fairfax, VA 22030

Lasse Holstrom
Center for Computational Statistics
George Mason University
Fairfax, VA 22030

David J. Cooke
Center for Computational Statistics
George Mason University
Fairfax, VA 22030

Kristine L. Bell
Applied & Engineering Statistics
George Mason University
Fairfax, VA 22030

Harry L. Van Trees
C31
George Mason University
Fairfax, VA 22030

Jock Grynovicki
3512 Parkfalls Dr.
Baltimore, MD 21236

Linda L. Hall
ATTN: STEAC-LI-B (Bldg. 359)
Aberdeen Proving Ground, MD 21005

Robert Hammell
U.S. Army Research Laboratory
ATTN: AMSRL-IS-CI
Aberdeen Proving Ground, MD
21005-5067

Tom Hettmansperger
Dept. of Stat., 317 Thomas Blvd.
Penn State University
University Park, PA 16802

USA ATC
ATTN: STEAC-EN-BA B. Kaschenbach
Aberdeen Proving Ground, MD 21005

Eric Lagergren
NIST
Bldg. 820 Rm. 353
Gaithersburg, MD 20899

Craig R. Morrisette
13025 Broadmore Rd.
Silver Spring, MD 20904

Linda Moss
U.S. Army Research Laboratory
ATTN: AMSRL-SL-BE
Aberdeen Proving Ground, MD 21005

David H. Olwell
173C Lee Rd.
West Point, NY 10996

Carey E. Priebe
Dept. of Mathematical Sciences
Johns Hopkins University
Baltimore, MD 21218

Dan Ralescu
Dept. of Mathematical Sciences
Univ. of Cincinnati
Cincinnati, OH 45221

Carl Russell
JNTF/SE
730 Irwin Ave.
Falcon AFB, CO 80912

Otto Schwalb
Dept. of Stats, P.O. Box 1892
Rice University
Houston, TX 77251

Jayaram Sethuraman
Dept. of Statistics
Florida State University
Tallahassee, FL 32306

Thomas R. Walker
185 Farm Rd.
Aberdeen, MD 21001

John R. Weaver
16007 Prairie Ronde
Scoolcraft, MI 49087

Daniel Willard
102 Army Pentagon
ATTN: SAUS-OR
Washington, DC 20310

David W. Hutchison
OCSA-PAGD
Pentagon
Washington, DC 20310

Michael D. McKay
Los Alamos National Laboratory
MS F600
Los Alamos, NM 87545

Lyle H. Ungar
Dept. of Computer & Information Science
University of Pennsylvania
Philadelphia, PA 19104

Denise Bullock
USAE Waterways Experiment Station:
CEWES-GM-K
3909 Halls Ferry Rd.
Vicksburg, MS 39180

Nancy A. Renfroe
USAE Waterways Experiment Station:
CEWES-GM-K
3909 Halls Ferry Rd.
Vicksburg, MS 39180

Douglas Tang
Walter Reed Army Inst. of Research
Washington, DC 20307

Daniel B. Carr
Center for Computational Statistics
George Mason University
Fairfax, VA 22030

Ngoc A. Le
8326 Highclifte Ct.
Annandale, VA 22003

Mark Vangel
Stat. Engineering Div.
820 W. Diamond Ave.
Gaithersburg, MD 20899

Kent M. Miller
4513 Foxwood Dr.
Chester, VA 23831

Dr. Lorrie Hoffman
Dept. of Statistics
University of Central Florida
Orlando, FL 32816

Turkan K. Gardenier
115 St. Andrews Dr.
Vienna, VA 22180

Deloris Testerman
1670 S. Via Cielo, Yuma, AZ 85364
U.S. Army Yuma Prov. Gr.; STEYP-RS
Yuma, AZ 85365

Scott B. Huxel
UNISYS Corp. Suite 300
1401 Wilson Blvd.
Arlington, VA 22209

Judea Pearl
University of California
Los Angeles, CA 90024

Robert L. Launer
Army Research Office
Math & Computer Sciences Division
P.O. Box 12211
Research Triangle Park, NC 27709

Harris Bernard
Dept. of Statistics
1210 W. Dayton St.
University of Wisconsin
Madison, WI 53706

Jeffrey L. Solka
Inst. for Computational Sciences
George Mason University
Fairfax, VA 22030

Wendy L. Poston
NSWC G-33
Dahlgren, VA 22448

Bradley C. Wallet
NSWC
Dahlgren, VA 22448

Eugene Dutoit
Dismounted Battlespace Battle Lab.
Fort Benning, GA 31905

INTENTIONALLY LEFT BLANK.

NO. OF COPIES ORGANIZATION

2 DEFENSE TECHNICAL INFORMATION CENTER
DTIC DDA
8725 JOHN J KINGMAN RD
STE 0944
FT BELVOIR VA 22060-6218

1 HQDA
DAMO FDQ
D SCHMIDT
400 ARMY PENTAGON
WASHINGTON DC 20310-0460

1 OSD
OUSD(A&T)/ODDDR&E(R)
R J TREW
THE PENTAGON
WASHINGTON DC 20301-7100

1 DPTY CG FOR RDE HQ
US ARMY MATERIEL CMD
AMCRD
MG CALDWELL
5001 EISENHOWER AVE
ALEXANDRIA VA 22333-0001

1 INST FOR ADVNCD TCHNLGY
THE UNIV OF TEXAS AT AUSTIN
PO BOX 202797
AUSTIN TX 78720-2797

1 DARPA
B KASPAR
3701 N FAIRFAX DR
ARLINGTON VA 22203-1714

1 NAVAL SURFACE WARFARE CTR
CODE B07 J PENNELLA
17320 DAHLGREN RD
BLDG 1470 RM 1101
DAHLGREN VA 22448-5100

1 US MILITARY ACADEMY
MATH SCI CTR OF EXCELLENCE
DEPT OF MATHEMATICAL SCI
MAJ M D PHILLIPS
THAYER HALL
WEST POINT NY 10996-1786

NO. OF COPIES ORGANIZATION

1 DIRECTOR
US ARMY RESEARCH LAB
AMSRL D
R W WHALIN
2800 POWDER MILL RD
ADELPHI MD 20783-1145

1 DIRECTOR
US ARMY RESEARCH LAB
AMSRL DD
J J ROCCHIO
2800 POWDER MILL RD
ADELPHI MD 20783-1145

1 DIRECTOR
US ARMY RESEARCH LAB
AMSRL CS AS (RECORDS MGMT)
2800 POWDER MILL RD
ADELPHI MD 20783-1145

3 DIRECTOR
US ARMY RESEARCH LAB
AMSRL CI LL
2800 POWDER MILL RD
ADELPHI MD 20783-1145

ABERDEEN PROVING GROUND

4 DIR USARL
AMSRL CI LP (305)

NO. OF COPIES	ORGANIZATION
1	HENRY C ALBERTS 5842 HILLDON ST MCLEAN VA 22101
1	KRAGG P KYSOR PO BOX 685 EDGEWOOD MD 21040
1	EDWARD J WEGMAN 10821 BURR OAK WAY BURKE VA 22015
1	GEORGE MASON UNIV PATRICIA K JOYCE CENTER FOR COMP STAT FAIRFAX VA 22030
1	DUKE UNIVERSITY DONALD A BERRY ISDS BOX 90251 DURHAM NC 27708
1	VA TECH JEFFREY B BIRCH DEPT OF STATS 210C HUTCHESON HALL BLACKSBURG VA 24061
1	BARNEY BISSINGER 281 W MAIN ST MIDDLETOWN PA 17057
5	WALTER REED ARMY INST OF RESEARCH ROBERT J BURGE DIV OF BIOMETRICS BLDG 83 WASHINGTON DC 20307
3	TEXAS TECH UNIV W J CONOVER COLLEGE OF BUS ADMIN LUBBOCK TX 79409
1	INDIANA UNIV OF PA D H FRANK 336 STRIGHT HALL INDIANA PA 15705

NO. OF COPIES	ORGANIZATION
1	HQ AFOTEC/CN FRANK B GRAY 8500 GIBSON BLVD SE KIRTLAND AFB NM 87117
1	JIM X CHEN 4212 SIDEBURN RD FAIRFAX VA 22030
1	JOHNS HOPKINS LAB/APL JOHN LYONS BLDG 24 RM E278 LAUREL MD 20723
1	LOS ALAMOS NATL LAB PAUL PEDERSEN MS B265 LOS ALAMOS NM 87545
1	DEFENSE INFO SYS AGENCY/D74 FRANCES J MORAWSKI 5600 COLUMBIA PIKE 5TH FLOOR FALLS CHURCH VA 22041
1	CTR FOR COMPUTATIONAL STAT ADI WILHELM GEORGE MASON UNIV FAIRFAX VA 22030
1	CINDY KRIEGER 107 BLUEBERRY CT STAFFORD VA 22554
1	RICE UNIVERSITY JAMES R THOMPSON DEPT OF STATISTICS 6100 S MAIN HOUSTON TX 77005
5	DAVID CRUESS 6123 CAMELBACK LANE COLUMBIA MD 21045
1	GEORGE MASON UNIV KWANG-SU YANG STATISTICS-CSI FAIRFAX VA 22030

NO. OF COPIES	ORGANIZATION
1	WRIGHT STATE UNIV THOMAS SUDKAMP DEPT OF COMP SCIENCE DAYTON OH 45435
1	EDWARD J MULROW 1061 N GEORGE MASON DR ARLINGTON VA 22205
1	GEORGE MASON UNIV JUERGEN SYMANZIK CTR FOR COMP STAT FAIRFAX VA 22030
1	GEORGE MASON UNIV LASSE HOLSTROM CTR FOR COMP STAT FAIRFAX VA 22030
1	GEORGE MASON UNIV DAVID J COOKE CTR FOR COMP STAT FAIRFAX VA 22030
1	GEORGE MASON UNIV KRISTINE L BELL APPLIED & ENGINEERING STAT FAIRFAX VA 22030
1	GEORGE MASON UNIV HARRY L VAN TREES C31 FAIRFAX VA 22030
5	JOCK GRYNOVICKI 3512 PARKFALLS DR BALTIMORE MD 21236
1	PENN STATE UNIV TOM HETTMANSPERGER DEPT OF STAT 317 THOMAS BLVD UNIVERSITY PARK PA 16802
1	NIST ERIC LAGERGREN BLDG 820 RM 353 GAITHERSBURG MD 20899

NO. OF COPIES	ORGANIZATION
1	CRAIG R MORRISETTE 13025 BROADMORE RD SILVER SPRING MD 20904
1	DAVID H OLWELL 173C LEE RD WEST POINT NY 10996
1	JOHNS HOPKINS UNIV CAREY E PRIEBE DEPT OF MATH SCIENCES BALTIMORE MD 21218
1	UNIV OF CINCINNATI DAN RALESCU DEPT OF MATH SCIENCES CINCINNATI OH 45221
5	JNTF/SE CARL RUSSELL 730 IRWIN AVE FALCON AFB CO 80912
1	RICE UNIVERSITY OTTO SCHWALB DEPT OF STATS PO BOX 1892 HOUSTON TX 77251
1	FLORIDA STATE UNIV JAYARAM SETHURAMAN DEPT OF STATISTICS TALLAHASSEE FL 32306
1	THOMAS R WALKER 185 FARM RD ABERDEEN MD 21001
1	JOHN R WEAVER 16007 PRAIRIE RONDE SCOOLCRAFT MI 49087
1	DANIEL WILLARD 102 ARMY PENTAGON SAUS-OR WASHINGTON DC 20310

NO. OF COPIES	ORGANIZATION
1	DAVID W HUTCHISON OCSA-PAGD PENTAGON WASHINGTON DC 20310
1	LANL MICHAEL D MCKAY MS F600 LOS ALAMOS NM 87545
1	LYLE H UNGAR DEPT OF COMP & INFO SCIENCE PHILADELPHIA PA 19104
2	USAE WATERWAYS EXPERIMENT STATION CEWES GM K DENISE BULLOCK NANCY A RENFROE 3909 HALLS FERRY RD VICKSBURG MS 39180
5	WALTER REED ARMY INST OF RESEARCH DOUGLAS TANG WASHINGTON DC 20307
1	GEORGE MASON UNIV DANIEL B CARR CTR FOR COMP STAT FAIRFAX VA 22030
1	NGOC A LE 8326 HIGHCLIFTE CT ANNANDALE VA 22003
5	MARK VANGEL STAT ENGINEERING DIV 820 W DIAMOND AVE GAITHERSBURG MD 20899
1	KENT M MILLER 4513 FOXWOOD DR CHESTER VA 23831
3	UNIV OF CENTRAL FLORIDA LORRIE HOFFMAN DEPT OF STATISTICS ORLANDO FL 32816

NO. OF COPIES	ORGANIZATION
1	TURKAN K GARDENIER 115 ST ANDREWS DR VIENNA VA 22180
5	US ARMY YUMA PROV GROUND STEYP RS DELORIS TESTERMAN 1670 S VIA CIELO YUMA AZ 85364
1	SCOTT B HUXEL UNIXYS CORP STE 300 1401 WILSON BLVD ARLINGTON VA 22209
1	UNIV OF CALIFORNIA JUDEA PEARL LOS ANGELES CA 90024
5	USA RSRCH OFC MATH & COMP ROBERT L LAUNER SCIENCES DIV PO BOX 12211 RESEARCH TRIANGLE PARK NC 27709
1	UNIV OF WISCONSIN HARRIS BERNARD DEPT OF STATISTICS 1210 W DAYTON ST MADISON WI 53706
1	GEORGE MASON UNIV JEFFREY L SOLKA INST FOR COMP SCIENCES FAIRFAX VA 22030
1	WENDY L POSTON NSWC G-33 DAHLGREN VA 22448
1	NSWC BRADLEY C WALLET DAHLGREN VA 22448
5	EUGENE DUTOIT DISMOUNTED BATTLESPACE BATTLE LAB FT BENNING GA 31905

NO. OF
COPIES ORGANIZATION

ABERDEEN PROVING GROUND

21 DIR USARL
AMSRL IS CI
BARRY BODT (10 CPS)
ANN E. M. BRODEEN
ROBERT HAMMELL
BARBARA D. BROOME (BLDG 116A)
AIVARS K. CELMIŅŠ (BLDG 116A)
AMSRL IS
JAGDISH CHANDRA
AMSRL SL BE
LINDA MOSS
AMSRL WM BC
DAVID W. WEBB (5 CPS)

2 DIR USAATC
STEAC LI B
LINDA L. HALL (BLDG 359)
STEAC EN BA
B. KASCHENBACH

1 DIR ERDEC
SCBRD RTM
KYONG PARK (BLDG E3330 RM 120)

INTENTIONALLY LEFT BLANK.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1998	Final, 22-24 Oct 97

4. TITLE AND SUBTITLE
Proceedings of the Third Annual U.S. Army Conference on Applied Statistics, 22-24 October 1997

5. FUNDING NUMBERS
AH80

6. AUTHOR(S)
Barry A. Bodt, Editor

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
ATTN: AMSRL-IS-CI
Aberdeen Proving Ground, MD 21005-5067

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-SR-74

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)

10.SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The third U.S. Army Conference on Applied Statistics was hosted by George Mason University (GMU) during 22–24 October 1997 at the recently opened Johnson Center on campus. The conference was cosponsored by the U.S. Army Research Laboratory (ARL), the U.S. Army Research Office (ARO), the U.S. Military Academy (USMA), the U.S. Army Training and Doctrine Command (TRADOC) Analysis Center-White Sands Missile Range, the Walter Reed Army Institute of Research (WRAIR), and the National Institute for Standards and Technology (NIST). The U.S. Army Conference on Applied Statistics is a forum for technical papers on new developments in statistical science and on the application of existing techniques to Army problems. This document is a compilation of available papers offered at the conference.

14. SUBJECT TERMS
applied statistics, experimental design, statistical inference

15. NUMBER OF PAGES
145

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

INTENTIONALLY LEFT BLANK.

USER EVALUATION SHEET/CHANGE OF ADDRESS

This Laboratory undertakes a continuing effort to improve the quality of the reports it publishes. Your comments/answers to the items/questions below will aid us in our efforts.

1. ARL Report Number/Author ARL-SR-74 (Bodt, Editor) Date of Report July 1998

2. Date Report Received _____

3. Does this report satisfy a need? (Comment on purpose, related project, or other area of interest for which the report will be used.) _____

4. Specifically, how is the report being used? (Information source, design data, procedure, source of ideas, etc.) _____

5. Has the information in this report led to any quantitative savings as far as man-hours or dollars saved, operating costs avoided, or efficiencies achieved, etc? If so, please elaborate. _____

6. General Comments. What do you think should be changed to improve future reports? (Indicate changes to organization, technical content, format, etc.) _____

CURRENT ADDRESS

_____
Organization

_____
Name E-mail Name

_____
Street or P.O. Box No.

_____
City, State, Zip Code

7. If indicating a Change of Address or Address Correction, please provide the Current or Correct address above and the Old or Incorrect address below.

OLD ADDRESS

_____
Organization

_____
Name

_____
Street or P.O. Box No.

_____
City, State, Zip Code

(Remove this sheet, fold as indicated, tape closed, and mail.)
(DO NOT STAPLE)